Univerzita Karlova v Praze
Filozofická fakulta
Ústav anglického jazyka a didaktiky

Název práce: Carrollova *Alenka v říši divů:* korpusově stylistická perspektiva

Title of Thesis: A Corpus Stylistic Perspective on Lewis Carroll's *Alice's Adventures in Wonderland*

DIPLOMOVÁ PRÁCE

Zpracovala: Jana Hrdličková, Bc.

Vedoucí práce: Mgr. Anna Čermáková, Ph.D.

Srpen 2015 Praha

Prohlašuji, že jsem diplomovou práci vypracovala samostatně a že jsem uvedla všechny použité prameny a literaturu. Souhlasím se zapůjčením diplomové práce ke studijním účelům.“

“I declare that the following MA thesis is my own work for which I used only the sources and literature mentioned. I have no objections to the MA thesis being borrowed and used for study purposes.”

V Praze dne 11.08.2015,

Jana Hrdličková

Acknowledgment

Tímto bych chtěla poděkovat doktorce Mgr. Anně Čermákové, Ph.D. za její pomoc, neuvěřitelnou pečlivost a hlavně trpělivost ve vedení této práce. Její rady byly pro sepsání práce neocenitelné.

Abstract

Úkolem této práce je korpusově stylistická analýza *Alenky v říši divů* (1865) a představení možnosti využití korpusových metod ke studii literárního díla. V teoretické části se práce zabývá pojmy styl, norma, *aktualizace, deviace* (odchylka) a *prominence* (Leech, 2008), které jsou klíčovými termíny ve stylistické analýze. Posléze se v textu zmiňují pojmy „klíčová slova" (*keywords*) a „n-gramy" (*clusters*), které jsou zkoumány dále v teoretické části. Je vysvětlen vztah klíčových slov k referenčnímu korpusu a aspekty délky a frekvence n-gramů v textu. Důležitým elementem korpusově zaměřené práce je udávané snížení zaujatosti badatele, který se v analýze zabývá statisticky doloženými markantními znaky díla.

V další části se práce zabývá pozicí *Alenky v říši divů* mezi dalšími příslušníky dětské literatury tzv. „Zlatého věku" (Golden Age of children's literature) (Knowles & Malmkjaer, 1995), rolí nesmyslu a narušování pragmatických pravidel interakce. Jsou popsány referenční korpusy a předloženy předběžné hypotézy.

V praktické části jsou nejprve zkoumána klíčová slova, která jsou rozdělena dle slovních druhů a následně svých sémantických kategorií. Tato slova jsou interpretována z hlediska jejich role v textu a vzájemné provázanosti, a je také zkoumán jejich kontext. Dále se práce zaměřuje na čtyř slovné n-gramy s nejnižším počtem výskytu čtyři, které následně blíže zkoumá. Poslední část je věnována pragmatice textu, při níž jsou využity Griceovy principy kooperativnosti (Grice, 1975) a Leechův princip zdvořilosti (Leech, 2008). Na závěr je zhodnocen přísun korpusového přístupu ve studii literárního díla.

The goal of the present thesis is a corpus stylistic analysis of *Alice in Wonderland* (1865), and it presents the possibilities of using corpus methods in the study of literary texts. In the theoretical part, the thesis is concerned with an explanation of the words norm, *foregrounding, deviance* and *prominence* (Leech, 2008), which are key terms in stylistics. The text then focuses in detail on various concerns connected to *keywords* and *clusters*, which are investigated later on in the theoretical part. The relation of keywords to the reference corpus and aspects of the length and frequency of occurrence of clusters are analyzed. An important element in a corpus-based analysis is that it is supposed to decrease bias, as the researcher studies statistically based significant markers of the text.

The next part of the thesis focuses on the position of *Alice in Wonderland* between other members of the children's literature of the Golden Age (Knowles & Malmkjaer, 1995), and the role of nonsense and non-observance of pragmatic principles of interaction in the text. The reference corpora are described and primary hypotheses stated.

In the practical part, keywords are studied first. They are divided into part of speech categories and interpreted from the point of view of their role and mutual relations in the text. Their semantics and context are studied as well. The thesis also focuses on four-word clusters with the minimal frequency of occurrence of four, which are further analyzed in detail. The last part is devoted to the pragmatic relations of the text, where Grice's Maxims (Grice, 1975) and the Maxims of the Cooperative Principle (Leech, 2008) are employed. In the conclusion, the benefits of the corpus approach to the study of literary texts are evaluated.

Contents

Tables:

Figures:

Excerpts:

Abbreviations:

AW – Alice in Wonderland
ACHcorp – keyword list of Alice as the study corpus and children's literature as the reference corpus
ACOcorp- keyword list of Alice as the study corpus and contemporary (Victorian) literature as the reference corpus

1. Introduction

Corpus stylistic analysis of literary texts is an approach which is becoming increasingly popular in linguistic and literary research. This is because it is a practical method that is based on objective, statistically-based analysis. Its strength also rests in the fact that it can help reveal previously unobserved patterns (Norgaard, et al., 2010: 4). The corpus approach should not be seen as mere data crunching, but as interplay between the quantitative data provided by software and a qualitative analysis performed by an individual researcher. While the corpus approach helps to reduce researcher bias (Fischer-Starcke, 2009: 500), an individual perspective cannot, and should not, be discounted (Hitt, 2012: 29).

In the corpus stylistic approach, the focus is especially on repeating patterns of language use (Fischer-Starcke, 2010: 34), the concordance lines in which they are used, and their collocates. The scope of an individual analysis may vary, but a more holistic approach is undertaken by those linguists who examine individual words as well as larger patterns and dispersion plots of a text. Since word strings are thought to have "more semantic and grammatical impact on meaning than single words" (Stubbs, 2007 cited in Fischer-Starcke, 2010: 113), an analysis of 'clusters' (see Section 2.5) facilitates a more rounded perspective. In the corpus approach, a 'keyword' (see Section 2.4) or a cluster list are generated by a software tool and the text's style markers are identified and analysed.

In the following study, I investigate Lewis Carroll's famous children's book, *Alice's Adventures in Wonderland (*1865*).* Its uniqueness lies not only in its pervasive linguistic play, but also in the fact that it represents one of the first of children's texts whose focus is on a child's individuality and the pleasure it derives through reading (Carpenter: 2009). The goal of the analysis is to demonstrate the usefulness of the corpus linguistic approach to the study of stylistics, and it will hopefully provide new insight into a much-loved work of the Golden Age of children's literature. For the analysis, the *WordSmith Tools* 6.0 (Scott, 2015) software will be used and two reference corpora, one comprising of children's literature of the Golden Age, the other of Victorian literary texts, will be employed (Section 4).

The study first presents the theoretical framework of stylistic and corpus linguistic concerns, such as the notion of norm, prominence, deviance or 'foregrounding' (e.g. Leech, 2008: 30)

(Section 2.1) and the corpus approach to style (Section 2.2). The works of prominent corpus linguists are discussed in Section 2.3, while key notions connected to the study of keywords and word clusters are touched upon in Sections 2.4 and 2.5. Section 3.1 discusses the software used for the analysis and the research method, while Section 3.2 is concerned with the choice of a reference corpus in the analysis of keywords and cluster length and cut-off point. In section 4, I discuss *Alice in Wonderland* and its connection to the Golden Age of children's literature and nonsense. At its conclusion, I list the pragmatic principles which will be used later on in the analysis. In the research part, I start with the smallest unit, the keyword (Sections 5.1 – 5.6), and continue through to larger, phraseological units (Section 5.7.1 – 5.7.8). Finally, the text's subversive pragmatics is discussed in selected passages in connection to Grice's maxims and the maxims of the Cooperative Principle (Section 5.8).

2. Theoretical background

2.1. What is style?

The question that needs to be asked at the outset of a study dealing with corpus stylistics is what is meant by the word *style*. Geoffrey Leech and Mick Short (2007: 9-10) describe it as language use, historically mainly associated with works of literature. According to them, style is "a relational term" (Leech & Short, 2007: 10) and linguists talk of it in connection with authors, historical periods, genres, and literary forms. They define the purpose of stylistics as a tool for explaining the connection between language and its artistic function (Leech & Short, 2007: 11). Another view of style is to see it as "variation of language use" (Norgaard et al., 2010: 155-156). Its exact definition is complicated by the notions of norm, purpose, context and authorship, and it is motivated by personal choices of a speaker and his/her "socio-cultural factors", which are governed by existing conventions.

When pursuing a textual analysis, the two key questions linguists put forward are how? and why?- that is, does a text have a particular effect on the reader and why did the author choose a specific expression (Leech & Short, 2007: 11)? These questions necessarily imply reference to a norm, which can be that of the given language, period or genre, etc. Any work concerned with style therefore has to choose a viable norm and then pursue those features which display the most salient textual prominence. Norgaard et al. importantly point to the fact that features are always context-dependent (Norgaard et al., 2010: 95). Leech and Short (2007: 41, 43) speak of norm in the same stance as of style; they define it as a relative concept. The closer it is to the chosen text, the bigger the probability that the features identified are really significant. Hanks (2013: 145-147) defines norm as "prototypes of usage, associated with prototypical syntactic behavior, which is associated in turn with prototypical beliefs about presupposition and entailment". An important characteristic of the norm is that its usage is "repeated" and widely socially accepted. As it changes with time and differs for different genres, it is important to choose a norm that is valid both for a specific period of time, and a specific genre.

The concrete features that linguists study differ considerably and include phonological patterns of a given literary work, individual words, word clusters, entire sentences or larger

chunks of text. These are generally assumed to have the greatest literary relevance. The schema below illustrates the relation between literary relevance, psychological prominence and statistical deviance. The technical term for prominence is 'foregrounding'.[1] It is a psychological phenomenon, and can also be called "effect". On its basis, a reader assigns relevance to statistically deviant features. It needs to be said, however, that relevance is by and large a subjective phenomenon. We presume on the basis of a marked deviance (a statistically countable difference from a given standard) that a feature is likely also to be relevant to the data in question. However, importantly, Leech (2008: 163, 176) points out that "not all deviation is explicable in terms of foregrounding", since we can posit a situation where deviant features do not become prominent and the rules for foregrounding therefore do not apply. That is because when talking of relevance and prominence, a linguist leaves the sphere of objectively quantifiable phenomena and enters that of individual appreciation, which necessarily differs for everyone. Mahlberg (2012: 8) adds that "[p]sychological effects are difficult (or impossible) to describe on the basic of corpus data alone". When using only corpus data as evidence, therefore, the researcher should refrain from making claims about presumed psychological prominence.

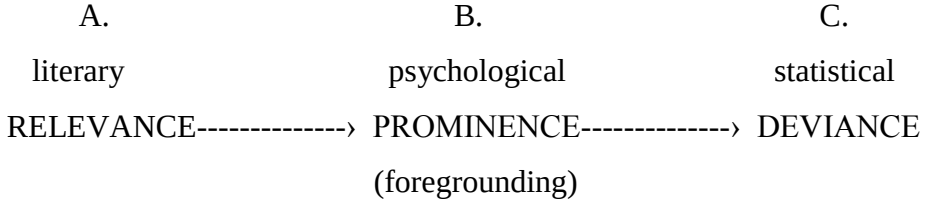


Figure 1: Leech & Short 2007: 41.

According to Leech and Short (2007: 39), the subjective effect of a text on a reader depends on a number of factors, such as the reader's "attentiveness, sensitivity to style and previous reading experience" which constitute our "linguistic competence". Psychologically prominent patterns[2] can be measured via reader response to specific chunks of texts, as was done for instance in Mahlberg, et al.'s "Reading Dickens's characters" (2014: 371). The study was based on Mahlberg's previous analysis of clusters in Dickens's fiction (Mahlberg, 2012). The researchers used eye-tracking to observe study participants while reading body clusters in

[1] Leech (2008: 18) states that "the Czech term aktualisace (…) used by the pre-war Prague School of linguistics" corresponds to his use of the term.

[2] Norgaard et al. (2010: 96) mention Emmot (1996, 2002a), Gibbs et al. (2002), Miall and Kuiken (1994) and especially van Peer (1986, 2002, 2007) as linguists studying the principle of psychological prominence in a text. According to Norgaard et al., van Peer affirms that foregrounded elements are processed more slowly and affective reader response increases.

Dickens and they discovered that the body language clusters were read significantly faster than other text extracts, and that the participant were even able to recall them when prompted. This was because, it is hypothesized, they were stored as units in long-term memory, and therefore easily recalled.

The concept of foregrounding concerns mainly parallelism and deviation (Norgaard et al., 2010: 95). In *Language in Literature* (Leech, 2008: 59-63), Leech considers three types of deviation. While he adduces that deviation especially characterizes poetic language (Leech, 2008: 59), it is clear that the definitions he gives are applicable to other forms of texts as well.

> **Primary deviation**: deviation from the norms of the language as a whole
> **Secondary deviation**: deviation from norms of literary composition
> **Tertiary deviation**: deviation from the norms internal to a text

Mahlberg (2012: 9) adds to this schema information relevant especially to corpus studies. Her delineation of the differences between the three types of deviation is described below.

1. **primary deviation** may be described by comparing a textual example to a general purpose corpus (i.e., a corpus that is taken as a sufficiently diverse sample of the language as a whole)
2. **secondary deviation** may be described by comparing a textual example to a corpus of all the works by the author
3. **tertiary deviation** may be described by comparing a textual example to the whole text from which it is taken

While it is possible to study deviation from various points of view, in this study I shall focus on only one of them: secondary deviation. I shall not describe the language of *Alice in Wonderland* in reference to language as a whole; I will employ two specifically chosen corpora (see Section 4.2).

2.2. Style and the corpus

The traditional, non-computer-assisted analysis of literary texts relies strongly on subjectivity. In recent years, however, it has become possible, thanks to the advancement of computer-based techniques and greater online availability of texts, to study a work of literature from the

corpus linguistic approach. The suggested benefits of this approach are greater objectivity of research, possibility of handling large quantities of data, and chance of identifying previously overlooked patterns (Fischer-Starcke, 2009: 492). If the purpose of the analysis is not to find information about language in general, but about a specific work of fiction, for instance, the amount of data becomes restricted (in size, genre, period, etc.). As a result, this form of research is increasingly gaining popularity among stylisticians and linguists. At the same time, many linguists believe in the joint benefits of both qualitative and quantitative approach (Busse, 2010: 37; Hitt, 2012: 29; Mahlberg, 2010: 292), which does not put the intuition of linguists out of play, but enables them to base their statements on empirically observable data.

According to Fischer-Starcke (2010: 3), "[c]orpus linguistics (…) assumes a correlation between the frequency of a pattern and its significance in the data." In the domain of corpus linguistics, a word or a collection of words and their textual background are studied, and repeated patterns are computed and analysed statistically with regard to their frequency and prominence in comparison to a reference corpus/corpora (see Section 3.2.1). These are assembled collections of texts functioning as the norm (see Section 2.1) and providing the researcher with the necessary benchmark. Individual words which are particularly prominent in a particular text are identifiable through the keyword method (Section 2.4).

Lexical keywords are thought to serve as indicators of "aboutness" of the text (Bondi & Scott, 2010: 62), whereas grammatical keywords point to its phraseological structure. Identification of these is a starting point of the analysis, since not all keywords displayed are necessarily of interest to the researcher. A more holistic approach is undertaken by linguists who focus on prominent clusters (also called lexical bundles or n-grams, see Section 2.5) of varying length. In her research of the stylistics of Victorian fiction, Mahlberg (2010: 297) calls them the "building blocks of fictional worlds". Corpus approach may in this way facilitate the identification and subsequent classification of prominent patterns. In connection, all these methods build up a complex view of textually prominent features which together constitute what we call style.

Unfortunately, the methods employed in corpus linguistics are in a way restricted, as there are limitations to what software can do. For instance, a program can show us the most frequent patterns, but an analysis of the infrequent, but still relevant, ones still depends on a traditional approach. Fischer-Starcke (2010: 6, 7) likewise mentions the possible loss of researcher

individuality or the impossibility of identifying such features as metaphors[3] as the shortcomings of a corpus approach. Furthermore, the choice of the reference corpora may directly influence what we see, and there are issues concerning imprecise tagging (if tagging is used), or the question of which statistical measure to adopt in the identification of textual keywords (see Leech, 2008: 164, 168; Archer, 2009). All of these and more will have to be resolved in the future of corpus linguistic research, for solutions here see Sections 3.1, 3.2).

2.3. Corpus stylistics in literary analysis: corpus research of literary texts

In her article "Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*", Fischer-Starcke (2009) states that "fiction texts have only rarely been analysed by corpus linguistic techniques". This view is shared by Mahlberg, (2007: 2) who says that "corpus linguists have only recently developed an interest in what may be called 'corpus stylistics'". While it is true that the corpus approach to literature is still fairly novel, there is now a number of linguists who investigate a single text or several texts of varying length from a corpus stylistic perspective. This study continues in the footsteps of several prominent corpus stylisticians and it therefore behooves us first to discuss their contributions to the field.

In "Conrad, concordance, collocation", Michael Stubbs (2005) analysed Conrad's *Heart of Darkness* and provided corpus evidence of Conrad's play with contrasts, vagueness and unreliable knowledge presented by the narrator. He also studied the effect of word distribution and phraseology in the novel. According to him, there has been "a lot of literary criticism […], but very little linguistic analysis of the novel" (Stubbs, 2014: 4).

In "Top keyword abridgements of short stories", Michael Toolan (2006) focused on Joyce's short story "Two Gallants" and the cohesion afforded by the top character keyword. According to him, such a keyword "creates a form of foregrounding" and points the reader's attention to the key developments in the story (Toolan, 2006: 181).

In "Keyness and the character-talk of Shakespeare's Romeo and Juliet", Jonathan Culpeper (2009) analysed the speeches of six characters whose utterances are prominent in the play by comparing each individual speech with that of the remaining five. He also analysed the keywords' dispersion plot and their semantic domains. As a result, he was able to make

[3] Headway is being made in the domain of metaphor corpus research, where semantic annotation plays a key role (metaphorical mapping- Deignan, 2005: 170).

stylistic differences not only between the key aspects of the characters' utterances, but also between their most common semantic areas.

In "Work in progress in Corpus stylistics", Geoffrey Leech (2008) performed first a traditional and subsequently a corpus analysis of Virginia Woolf's "Mark on the Wall". In the traditional analysis, he was able to focus on slightly different features (e.g. phonetic effects or the functional sentence perspective of sentences), while the corpus analysis both provided confirmation for some of his earlier statements and pointed his attention to aspects which he had disregarded.[4] The corpus analysis was done by comparing the "Mark on the Wall" with a reference corpus of the 1890s novels and the 1930 general fiction and focused on the story's most frequent positive and negative keywords, parts of speech and semantic tags.

In "Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*", Fischer- Starcke (2009) examined the novel's keywords and 4-grams, particularly focusing on semantic patterns such as family relationships. A year later, in her book *Corpus Linguistics in Literary Analysis* (2010), she analysed Jane Austen's *Northanger Abbey* in comparison with a reference corpus of contemporary literature and a corpus of assembled Austen novels. She focused on keywords and concordance lines, phraseology and text segmentation. Keyword analysis enabled her to discuss key topical concerns of the novel's style, for instance the means of characterization or the role of various verb types, along with irony and negation. Four-word clusters she generated pointed to the novel's recurrent phrases, for instance those used for temporal and spatial orientation, those connected with personal pronouns, etc. Lastly, text segmentation was discussed in connection to cohesion and coherence in the novel on the basis of the novel's key recurrent lexis.

A seminal work in the domain of corpus stylistics with specific focus on lexical clusters was written by Michaela Mahlberg (2012), who analysed multi-word clusters in Dickens' fiction to gain further insight into his character description and speech patterns, their semantic roles and distribution: for instance the role of character speech and body language in the novel. In the study, the clusters are divided into five main semantic groups and each of them is discussed in detail.

[4] In the course of the analysis, he discovered (among other findings) a use of the generic pronoun *one* of which he had not previously been aware (Leech, 2008: 162-177).

2.4. Keywords

Keywords are one of the most frequently used methods in corpus analysis. The term denotes words which "occur statistically significantly more frequently in a text or corpus than in a comparable, larger reference text or corpus" (Fischer-Starcke, 2010: 65; Scott & Tribble, 2006). For the study of individually/researcher-compiled corpora, software such as Laurence Anthony's AntConc (e.g. version 3.4.3: Anthony, 2014) or Mike Scott's WordSmith (e.g. version 6: Scott, 2015) can be employed. With aid of these, a researcher is able to generate a keyword or a cluster list together with measures of frequencies or keyness values according to a specific probability measure. Frequency and keyness lists provide exact frequency and statistical information needed for further qualitative analysis. The method of keyword calculation is a matter of debate, but commonly the choice lies between two camps: chi-square and the log likelihood measures.[5] The chi-square method "becomes unreliable when the expected frequency is too small" (Rayson, et al., 2004: 3), which is one of the reasons why many linguists argue against its use as the results may become skewed[6] (Rayson et al., 2004: 4). Another important concept in keyword analysis is statistical probability. WordSmith's default setting is 0,0000001000000012 using the log likelihood method. This is for reasons of selectivity, and because it enables the researcher to generate a manageable amount of keywords.

Many linguists are of the opinion that open-class words are the text's chief indicators of "aboutness", while closed-class keywords point rather to the stylistic markers of the text (Scott, 2000 cited in Bondi & Scott, 2010: 62; also Baker, 2004). Text aboutness relates to the plot of the text and its main concerns. On the other hand, closed-class words can point, for instance, to the role of negation in the text (e.g. in Fischer-Starcke, 2010) or indicate words which are part of larger, significant textual units (e.g. the correlation of 'if' and 'be' in Culpeper's (2009) analysis of *Romeo and Juliet*.[7] Depending on their dispersion throughout the text, they form plots which indicate where in the text a specific keyword clusters.

[5] See also Gabrielatos, C., & Marchi, A. (2012). *Keyness: Appropriate Metrics and Practical Issues*. Retrieved from http://repository.edgehill.ac.uk/4196/1/Gabrielatos%26Marchi-Keyness-CADS2012.pdf and Hardie, A. (2014). *Log Ratio: An Informal Introduction.* Retrieved from http://cass.lancs.ac.uk/?p=1133.

[6] In their article "Keyness: Matching metrics to definitions" (2011), Gabrielatos and Marchi propose a new metric for the calculation of keyness to avoid the effects of sample size: they suggest measuring frequency difference.

[7] In his "Keyness and the character talk of Shakespeare's Romeo and Juliet, Jonathan Culpeper suggests a division of keywords among 'ideational', 'textual' and 'interpersonal' (Culpeper, 2009).

Another possible concern for linguists are negative keywords which appear in a text significantly *less* frequently than in a given reference corpus. They are computed because statistical deviance is not the property only of prominently present items, but also of prominently absent ones. Negative keywords may indicate the low frequency of certain items in the text as a significant style marker. The presence of negative keywords is based on sufficient disparity of the reference corpora from the study corpus and establishes the text's markedly under-represented items. Leech (2008: 171) gives an example of negative keywords in "The Mark on the Wall", for instance *he, she* or *you* and their possessive and demonstrative variants. These words are conspicuously absent from the story since it has a first person narrator who addresses no-one. A similar analysis was done by Culpeper (2009: 38), who identified *Romeo* as a negative keyword in Romeo's speech, pointing to the fact that Romeo does not habitually self-reference.

Sorting keywords by semantic content provides valuable information regarding the aboutness of the text, although in a wider sense they can also serve as indicators of genres (Baker, 2004). Assigning keywords to semantic categories provides the researcher with an overview of the text whose semantic domains might remain opaque in the course of a traditional stylistic analysis. A number of linguists have used semantic analysis in order to gain information about major thematic concerns of a given text. This can be done either manually or by means of a tagging software. Fischer-Starcke (2009: 501) manually identified 'family and family relationships' as one of the prominent topics of Jane Austen's *Pride and Prejudice*, while Jonathan Culpeper (2009) used the SEMTAG program in order to discuss the key semantic groups in *Romeo and Juliet*. A similar software-assisted analysis was performed by Leech (2008: 174) on Virginia Woolf's "Mark on the Wall". While software makes the work easier, manual sorting has its advantages, especially since there is less danger of mistakenly tagged items. On the other hand, manual sorting can be difficult and time consuming in the case of larger corpora. In the present analysis, manual sorting will be employed.

2.5. Clusters

Clusters (n-grams) are phrasal units of varying scope, which are becoming a widely used methodology in the corpus study of texts (see e.g. Fischer-Starcke, 2009, 2010; Mahlberg, 2007, 2012). The reason for the study of clusters is mainly the belief that meaning is not encoded in solitary words, but rather in larger phrasal units (Sinclair, 1991). According to

Stubbs (2007, cited in Starcke 2010: 113) "word strings have more semantic and grammatical impact on meaning than single words." Mahlberg (2012: 50) adduces that the study of phraseological units provides valuable information about their discourse functions and about "prefabricated phrases" which are stored in our mental lexicon. Starcke (2010: 108) states three main objectives of cluster researchers:

a. to draw conclusions about frequent or dominant contents of the data based on the lexis which occurs in phrases
b. to analyse recurrent words or phrases as cohesive links in the data
c. to identify frequent lexical and grammatical patterns (collocations and colligations) within and in the co-text of the word chains

She further points to two main focal points of the researchers' analyses, which illustrate the importance of clusters in corpus research (Starcke, 2010: 113):

a) the use of word-strings as meaning-encoding features
b) the function of word strings for textual organization

There are several terms used for what I here term as clusters, following Mahlberg's (2007, 2012) terminology, as her texts have been used here as key source. Mahlberg (2012: 52) defines clusters as "uninterrupted sequences of variable length, which stop at punctuation". Biber (2006) terms longer textual chunks as 'lexical bundles' and sees them as units "widely distributed across several texts", where centered bundles reflect the thematic units of discourse (Mahlberg, 2012: 51). The minimum frequency of occurrence, their length, the minimum number of texts in which they occur and the requirement that they do not cross punctuation are all necessary criteria for these multi-word units to be recognized as lexical bundles (Mahlberg, 2012: 52-52). Fischer-Starcke (2010) uses the terms 'n-grams' (uninterrupted sequences) and 'p-frames' (those that can be variable in one place).[8] N-grams are the most general term used in computational linguistics.[9]

[8] See Fischer-Starcke (2010: 108-109). According to her analysis, the most frequent 4-gram in *Northanger Abbey* is *i am sure i*, whereas the most frequent 4-frame is *the * of the*.

[9] There are also other terms, such as 'concgram'. According to Scott (2010) at <http://www.lexically.net/downloads/version5/HTML/index.html?definition_of_a_concgram.htm>, "the concgram procedure takes a whole corpus of text and finds all sorts of combinations (…), whether consecutive or not." The ConcGram programme for the identification of these combinations was first developed by Greaves

The focus of researchers who study clusters necessarily varies as well, from focus on spoken English (Altenberg, 1998) to functional and grammatical characteristics of clusters in different text types from the diachronic point of view (Culpeper & Kytö, 2002), or their variance in different text types in general[10] (Stubbs & Barth, 2003; Biber et al., 2004). In *Corpus Stylistics and Dickens' Fiction* (2012: 297), Mahlberg focuses on clusters indicative of 'local textual functions'.[11] She describes them as "textual building blocks" which contribute to a more holistic description of a work of literature, specifically as related to character creation. She is limited to one author only (as opposed to Biber or Culpeper and Kytö, who study clusters in various texts). Whereas a linguist who analyses various disparate texts can make wider statements about language (of a given genre, period, etc.) as a whole, the scope of a linguist who studies clusters in the work(s) of one author is relatively narrow. According to Mahlberg (2012: 53-54, 61), there are two other key characteristics of clusters; the fact that in fiction, there seem to be fewer lexical clusters than in other types of discourse, and that the more words a cluster contains, the more "text-specific" it is.

Mahlberg (2012) classifies clusters in Dickens's fiction as 'labels', 'speech clusters', 'body part clusters', 'as if clusters', 'time and place clusters' and "other". Fischer-Starcke (2010: 118, 119), on the other hand, identifies clusters expressing "temporal, spatial and/or quantitative/qualitative relationships" and those having a personal pronoun, modal verb, negation, or verbs/adjectives "describing perception of a mental process" at their core in her study of *Northanger Abbey*. An immediately noticeable point is that in any literary analysis, there must necessarily be large overlap in the semantic characteristics of the data. Fiction necessarily revolves around various expressions concerning time, space, number, quality, or the personal perceptions of the characters, including their speech. On the other hand, clusters most revelatory about the text are arguably those which are yet more specific, like Mahlberg's "body part" or "as if" clusters. Of course, there is also the possibility that a specific cluster group appears in one part of the text only. As a result, an analysis of the cluster distribution in the text is crucial for its holistic study.

(2005) and on the basis of it, WSConcGram was implemented by Scott in WordSmith. Another term is 'skipgram', which is used for non-contiguous and contiguous word associations alike (Cheng et al., 2006: 412).

[10] Cited in Mahlberg, 2007: 5.

[11] According to Mahlberg (2007: 4) "local textual functions are 'textual' as they describe the functions of words (or combinations of words) in text, and they are 'local' in that they do not claim to capture general functions, but functions specific to a (group of) text(s) and/or specific to a (group of) lexical item(s)."

3. Methodology

3.1. The software and research method

The software used for the present analysis is version 6,0 of Mike Scott's *WordSmith Tools* (2015). The software's functions of Concord, Keywords and Wordlist shall all be employed here. With the Concord tool, the researcher is able to upload a text file and examine its concordance lines, clusters surrounding a specific node (focus) word, examine the text's clusters according to their number or distance from the node word and calculate the plot of given words. A 'concordance' is a set of lines including the search word plus its immediate context. When researching keywords, the context in which they are found is of paramount importance, as it can offer valuable clues as to the semantic colouring of the keyword and the stylistic layout of the text.

The purpose of the WordList tool is to create a list of words present in a chosen corpus or corpora. These lists can subsequently be used in the KeyWords tool where a comparison between them is made. Word frequencies and percentages in the study and the reference corpora can be computed, as well as the words' keyness value at a given probability level. In the present analysis, the default probability level of 0,0000001000000012 shall be used.

The 'links' tool serves the purpose of listing the key collocates of a given keyword (the default setting being 1 to 5 to the left and the right side of the node), the link types and the number of keyword hits. Keyword plot along with a measure of its dispersion, keyness, link tokens and hits can also be computed. Together, these functions provide a wide statistically based research ground of a given text.

'Collocation'[12] denotes a habitual co-occurrence of structures (see discussion of 'norm' in Section 2.1) which lends the words a "semantic colouring" and serves to create meaning generally opaque to human intuition (Wynne, 2006: 225). Sinclair (2004: 14) defines it as "the co-occurrence of words with no more than four intervening words", due to their psychological association (cited in Norgaard et al., 2004: 57). For this reason the collocation

[12] The term 'colligation' is used for a similar principle of association between grammatical, as opposed to lexical, items (Norgaard et al., 2010: 57). See also "Collocation, colligation, semantic prosody and semantic preference of the core word" (Sinclair, *The Lexical Item* (1998)) in E. Weigand (Ed.), *Contrastive Lexical Semantics (Current Issues in Linguistic Theory*, 1998.)

span (or horizon) is usually that of five words to the left of the node word and five words to the right. Olohan (2004: 82) speaks of it as a commonly employed setting. WordSmith's default setting likewise shows a collocate span of 5L – 5R, where concrete collocates are displayed and their frequencies given. WordSmith also allows the computation of relationship strength between the node word and its collocate (the default statistical method being the MI score)[13]. This provides researchers with statistical data with which to support their claims about the significance of concrete collocates.

A connotation is a term denoting a word's associated meaning, based on cultural, individual, and historical experience of the speaker and the listener.[14] Fischer-Starcke (2010) used connotations in order to examine in detail the context of keywords in her analysis of *Northanger Abbey* (e.g. the negative connotations of the word 'novel', 'journal' and 'manuscript' or the period-reflecting connotations of 'family')[15] and a similar process was done by Stubbs in his study of *Heart of Darkness* (e.g. the death connotations of the word 'grass' or the atmosphere of danger connected with the words 'gleam' and 'glitter').[16] Text segmentation can also yield profitable results. Fischer-Starcke (2010) analysed the text segmentation of *Northanger Abbey* and she found out that it deviated from the traditionally accepted pattern (e.g. place names, which continue to crop up even when the plot moves away from them).[17] She also analysed the places of occurrence of the novel's dominant lexis. Similarly, Culpeper (2009: 40-41) analysed the dispersion of Romeo's keywords in *Romeo and Juliet* to see which words refer to Juliet and which to Rosaline in the play.

The focus of this study will firstly be set on keywords (Sections 5.1 to 5.6). In order to understand these, they will be divided according to their part of speech. On their basis, the primary semantic notions of the text will be identified and discussed. I shall thereby gain valuable insight into *AW*'s thematic concerns and stylistic devices. Where profitable, the keywords' concordance lines shall be examined in detail and their collocations analysed. I will likewise look at the text's keyword plot and segmentation to examine where specific keywords cluster in the text. WordSmith sorts them by order of occurrence in the text, which helps in the identification of thematic structures of the text as they crop up. After the analysis

[13] More computational methods can be used in the programme, such as log likelihood, T-score, Z-score, specific mutual information or MI3.
[14] See Leech (1974: 40-41).
[15] Fischer-Starcke, 2010: 78.
[16] Stubbs, 2005: 14-15.
[17] Fischer-Starcke, 2010: 154.

of keywords, the text's four-word clusters shall be examined (Sections 5.7.1 to 5.7.8). According to Mahlberg (2012: 61), "[w]hen clusters increase in length they become more text-specific." She herself studied five-word clusters in Dickens's fiction and found this restriction profitable. As *AW* is a much shorter text and Carroll is less repetitive than Dickens, the generation of five-word clusters with minimal frequency of three occurrences yields only 28 clusters, which is a relatively low number. The frequency of three is already a very low threshold and should not be further decreased if the analysis is to provide any profitable results. For these reasons, we shall focus here on four-word clusters and the frequency threshold will be set to four (for more details see Section 3.2.2). Lastly, as nonsense interactions based on a non-observance of Grice's (1975) conversational maxims and Leech's (2008) politeness maxims play a great role in *AW*, I shall analyse selected Wonderland dialogue and demonstrate how the exchanges contribute to the non-productivity of conversation in the text (Section 5.8).

3.2. Keywords and Clusters

3.2.1. Choosing a reference corpus in keyword analysis

As every reference corpus will generate a slightly different keyword list, its choice is likewise of primary importance. However, as Teubert and Čermáková (2007: 69) state, there is "no standard recipe for the composition of a special corpus." In Fischer-Starcke's (2010: 66) research, the considerations for the choice of reference corpora were its larger size (than the study corpus), diachronic comparability and the equality of the text type. She also suggests the use of at least two reference corpora so that the high significance of items on both lists is not accidental (Fischer-Starcke, 2010: 65; Archer: 2009, 12).

Leech (2008: 166) likewise recommends the choice of a reference corpus from a comparable period of time and suggests a "scatter-gun approach", where the researcher performs a series of comparisons with various relevant norms, "preferably with reference corpora of different level generality" (Leech, 2008: 167). He adds that "*slight* differences of period, genre and so on among alternative reference corpora are unlikely to make substantive differences to the results" (Leech, 2008: 167) and therefore can be employed to advantage. This view is shared by Berber Sardinha (2004), who writes that the choice of similar texts in the reference corpus filters out common features (Archer, 2009: 8), and therefore certain heterogeneity should be observed.

In his "In Search of a Bad Reference Corpus", Scott (2006: 1) names "size in tokens, similarity of text-type, similarity of historical period, and similarity of subject-matter" as key concerns. What he discovered in the course of his analysis is that there is no "clear and obvious threshold below which poor keyword results can be expected" (Scott, 2006: 8), but that in general, the larger the reference corpus the better. However, large reference corpora may yield unmanageable amount of keywords for qualitative analysis. A deliberate choice of an incomparable reference corpus did not lead to inappropriate results, although more keywords were generated. Similarity of text-type was therefore found not to be strictly necessary. Although results differed when genre-different corpora were used, they were not conclusive. What can be gleaned from these differing views is that common sense should be used. While a deliberately unusual reference corpus might not devalue the results of the analysis, its larger size, a slight but not complete similarity of genre and an equality of text type and period are probably advisable.

There is likewise a considerable debate over the cut-off point of a keyword list. Even with a comparatively low p value threshold, the keyword list may become unmanageably long, which also ties in with the size of the reference corpus. One of the solutions is to dismiss some items which seem less promising in terms of being revelatory of the text's style, such as for instance proper names or various grammatical categories. While a researcher may be justified in doing so, there is always the possibility that the less likely items could prove to be significant upon more detailed analysis. Scott writes that "a text about racing could wrongly identify as key, names of horses which are quite incidental to the story" (cited in Culpeper, 2009: 38). Conversely, Michael Toolan (2006) uses proper names in James Joyce's short story analysis „Two Gallants" to prove their centrality to the narrative.

In their discussion of keyness, Gabrielatos and Marchi (2011: 2) state that "the vast majority of studies examine a subset of keywords – almost always the top 100 keywords as ranked by the metric used". In "Querying keywords", Paul Baker (2004) adduces that no consensus has been reached in the question of cut-off points. As this is the case, the decision rests with the informed judgment of any individual researcher and constitutes the qualitative part of the analysis. In the present study, as the number of keywords did not reach much above 100 (111 positive keywords in the corpus of contemporary literature and 72 in the children's literature corpus), a decision was made to include all in the research part.

3.2.2. Cluster length and cut-off point

In her study of *Northanger Abbey*, Fischer-Starcke (2010) focuses on four-word structures. Her explanation for this choice is that several linguists affirm that three-word structures are both frequent and simply collocational in nature, whereas longer structures' phrasal nature makes them "more specific to the data" (Fischer-Starcke, 2010: 111). Focusing on four-word structures is a compromise between these two extremes. Mahlberg (2012: 61) chooses five-word clusters that do not stop at punctuation in her study of Dickens's language, as "they appear to be sufficiently long to find text-specific features and at the same time provide sufficient data". However, she also admits that the choice is somewhat arbitrary (which is made possible by the repetitive nature of Dickens's texts).

Mahlberg (2012: 61) points out that for the study of fiction, the cut-off points of large corpus studies do not apply. She starts by focusing on a lower number of key clusters of the p-value of 0.00001, taking into account positive clusters only. The key comparison, according to her, is useful for finding a limited number of clusters in order to establish the cluster categories into which other frequent clusters can subsequently be sorted (Mahlberg, 2012: 64). As she is concerned with the entire corpus of Dickens's novels, investigation of key clusters appears necessary. For the study of a shorter text, such as *Alice in Wonderland*, the criteria employed may be less restrictive. Archer (2009: 59) likewise discusses the problem of the cut-off point, stating that "n-gram tables become quite unmanageable in terms of size" and that one of the solutions is "to include just those n-grams that occur three times or more". However, this potentially robs the researcher of the chance to investigate less common, but all the more interesting patterns. Since there is no given rule, it is perhaps best that all researchers make their own informed decision, which in this case is to focus on four-word clusters and include all (a total of 55) in the analysis.

4. The data

4.1. Alice's Adventures in Wonderland

4.1.1. Alice in Wonderland and the Golden Age of children's literature

In *Language and Control in Children's Literature*, Murray Knowles and Kirsten Malmkjaer posit the 1860s as the beginning of the "First Golden Age" of children's literature (Knowles & Malmkjaer, 1995: 16; also Hunt, 1994). This period starts with Carroll's *Alice in Wonderland* (the abbreviation *AW* will be used henceforth) and ends in the early 1900s (Knowles & Malmkjaer, 1995: 18). While books for children were written before Carroll as well, their primary purpose was moral education. They were therefore written for the purpose of instruction rather than solely enjoyment and they saw a child as a miniature adult. Into this field can be grouped traditional fairy tales, whose aim is to present their audience with a moral. As opposed to fantasy literature, into which *AW* is oftenest grouped, fairy tales' roots lie in the past and not the present, and they lack the three hallmarks of fantasy literature: an anchoring in reality, a protagonist experiencing spiritual maturation, and time distortion (Nikolajeva, 2003: 140-142), all of which happen in *AW* (as we shall see later on).

In creating *AW,* Carroll's aim was diametrically different from that of an instructional narrative; he used the medium of fantasy and nonsense to subvert traditionally expected rules of narrative along with expectations regarding the children's books genre. In fact, some critics claim that *AW* signals "liberty and thought in children's books" (Carpenter, 2009: 68) and that at its core, it contains elements which rather push it away from the domain of books for children.[18] This new movement was possible only with the advent of Romanticism and later on in the Victorian era, where the traditional view of children changed. The society stopped viewing them as future adults in need of moral and spiritual instruction, and started seeing them as individuals possessing their own unique perspective (Carpenter, 2009: 9). Furthermore, books for children often offered their authors a means of escape into the "nostalgic space of wild childhood", reconnecting them with their "idyllic past" (Leach & Hollingsworth, 2009: 23, 34). As such, they served the double purpose of entertaining the child and adult alike. However, *AW* does not entirely fit this description either. While

[18] Humphrey Carpenter writes in his *Secret Gardens: A Study of the Golden Age of Children's literature* that *Alice* is in fact "an exploration of violence, death, and Nothingness" as well as "a mockery of Christian belief", and that these themes are pursued through the medium of nonsense (Carpenter, 2009: 62).

Wonderland is a sort of remote, at times even idyllic place, the linguistic nonsense as well as the constant play with basic realities lends it a dark edge of permanent inconstancy which can be frightening. There are critics who term *AW*'s adventures as "a frightful journey into meaningless night" (Rackin, 1966 cited in Kincaid, 1973: 92).

As *AW* belongs in the domain of children's literature of the Golden Age, it might be profitable to group some of the books which represent it into broad categories which help to posit *AW* in the midst of their variety. All of the texts necessarily vary not only in their topic, character types and stylistic features, but also in the very psychology with which they were written. Knowles and Malmkjaer's *Language and Control in Children's Literature* (1995) and Carpenter's *Secret Gardens* (2009) were used as key sources. The titles discussed here form the corpus of children's literature used for the present analysis (see Section 4.2). They can be grouped into:

1. traditional juvenile fiction
2. imaginative fantasy literature
3. the fairytale
4. the Arcadian movement

The position of *AW* is under imaginative fantasy literature. Knowles and Malmkjaer (1995: 16-17) classify the *Alice in Wonderland* and *Beyond the Looking-Glass* books (1865, 1871) and Kingsley's *The Water Babies* (1863) in this category because the aim of these texts is to rebel against existing authority and tradition. Kipling's *Jungle Book* (1894) and Stevenson's *Treasure Island* (1883), on the other hand, are typical representatives of "traditional juvenile fiction", where the concerns of child character development, morality and the Empire assume primary importance (Knowles & Malmkjaer, 1995: 11-13). Macdonald's *The Princess and the Goblin* (1872) is seen as a fairytale, but one which is inherently subversive (Knowles & Malmkjaer, 1995: 164). Potter's *The Tale of Peter Rabbit* (1902) is described in Carpenter's *Secret Gardens* (2009) as another representative of the fairytale, this time with a predator-victim motive (Carpenter, 2009: 147). The book to which Carroll's *AW* probably comes the closest is Barrie's *Peter and Wendy* (1911), as both these books contain deeply subversive elements and offer a more or less realistic view of children operating in fantasy spaces (Carpenter, 2009: 182). On the other hand, Burnett's *The Secret Garden* (1911) is seen as a deeply Victorian, escapist text whose main topic is the search for the lost Arcadia (Carpenter,

2009: 182). Milne's *Pooh* (1926), according to Carpenter, is a book that can be fully appreciated by a child's mind, portraying the basic characters of humanity in its animal protagonists (Carpenter, 2009: 205).

4.1.2. Alice and Nonsense

In *Language in Literature*, Leech (2008: 56) talks of the different kinds of literary language associated with texts with varying purpose, such as 'scientific language', 'advertising language', etc. Linguists are able to make such distinctions because every type of language has its inherent characteristics which it shares with other members of its class. In the case of *AW*, language of nonsense immediately comes to mind. The term 'nonsense' can be misleading, however, because it implies an exclusion of sense. If applied literally, the resultant text would be so deviant from the language as a whole that it would be rendered incomprehensible.[19] The nonsense of Carroll (often written with a capital *N*) is diametrically different, as its effect rests on the fact that it is strictly subjected to rules. In fact, for nonsense to function systematically in a text, it has to be a "consciously regulated pattern" (Flescher, 1969: 128).[20] In other words, rules applying to everyday language still apply but are "reversed", and a word's relation with sense is strengthened and weakened by turns in order to endow it with a new, literal or non-literal, meaning (Flescher, 1969: 128, 134). As the text of *AW* often functions on the principle of intentional ambiguity of form and meaning, the work of a corpus linguist is doubly complicated. One of the subjective elements that need to be brought into the analysis is the disambiguation of formally indistinguishable patterns.

As Carpenter (2009: 55) writes, Carroll himself was not the inventor of nonsense, but very likely took his inspiration from Edward Lear's *Book of Nonsense* (1846). The reasons for why he did so must have been because of the potential that linguistic playfulness has for a child's mind. Fiona McArthur (2004: 53) sees the linguistic play in *Alice* as a sort of "dynamic problem-solving activity" while A. L. Lucas (1999: 159) speaks of the potential for rebellion and self expression every child needs manifested in the language of Wonderland. The role of nonsense in *AW* is therefore to provide its readers with the possibility to be playful and individual along with the protagonist and it presents a child's maturing mind such as it is, puzzled by the frequently arbitrary rules of behaviour and language of the adult world. By

[19] See Leech's concept of primary deviation (Leech, 2008: 59-63) discussed in Section 2.1.

[20] In "Alice's Invasion of Wonderland" James R. Kincaid writes that according to Elizabeth Sewell's study of *Alice*, the book's nonsense is not a frightening element, but rather something "deeply reassuring". This is because of an appearance of disorder where actually, order is strictly maintained (Kincaid, 1973: 92).

using nonsense holistically in a children's story, Carroll moved his literary creation one step closer to modernity and to what readers look for in a children's book today.

Nonsense in Wonderland does not function solely on word level: its dialogue and the rules by which it abides fall under its influence as well. Conversation is a game between the participants and becomes increasingly difficult to sustain as the meaning and basic pragmatic principles of interaction crumble. The pragmatic principles by which an ordinary dialogue is driven presume that both participants wish for their interaction to be maximally fortuitous. Needless to say, this is not the case in Wonderland. The rules by which a participant should ideally abide are best summed up in the conversational maxims of the philosopher H. P. Grice, given in the figure below[21]:

(i) *The maxim of quantity*
Give the required amount of information – not too much or too little.
(ii) *The maxim of quality*
Do not say that for which you lack evidence or which you believe to be false.
(iii) *The maxim of relation*
Make your contributions relevant to the purpose in hand.
(iv) *The maxim of manner*
Avoid obscurity, ambiguity and unnecessary prolixity, and be orderly.
Figure 2: Leech & Short, 2007: 236.

It is, of course, impossible for all the maxims to apply at once in conversation. In fact, Norgaard et al. (2010: 69) write that "it is a very common occurrence that interactants in conversations 'fail' to abide by such principles which are not to be seen as rules but rather as general expectations that underline communication." However, if almost no principles are kept, interaction becomes pragmatically untenable. There are two types of maxim non-observance. One of them is flouting, a "blatant and intentional non-observance of any of the maxims," while the other one is violation, where the speaker is aware of the pragmatic implicature created, but is neither ostentatious nor blatant (Norgaard et al., 2010: 69).

[21] There are, of course, more politeness models, for instance Brown and Levinson's (1987: 62) model of the positive face (taking into account the desirability of individual wants for others) and the negative face (the self's desires take precedence) (Norgaard et al., 2010: 136). Negative politeness includes maxims such as "be indirect", "do not presume or assume", "do not coerce your hearer", or "communicate your wants not to impinge upon the hearer's" (Norgaard et al., 2010: 138).

Norgaard et al. (2010: 39) state that pragmatic stylistics "contributes to the characterization of the protagonists" and aids in the creation of "power structures". It is an inherently interactive discipline and as such has an important place in the analysis of literary texts. The exchange of information taking place between characters on the basic level happens between the author and the audience on another (Norgaard et al., 2010: 41). The norm for a pragmatic analysis of a fictional text is therefore, of course, "authentic everyday communication" (Norgaard et al., 2010: 41), and it is this which enables the reader to make judgments about the pragmatic nature of a text.

A literary text entirely made up of pragmatic communicative interactions is drama. Drama holistically exploits the basic principles of communication in order to present an example of a reality-based social interaction. In *Language in Literature*, Leech (2008) focuses on the pragmatics of an absurd play, G.B. Shaw's *You Never can Tell*. His argument is that the principles of deviation are highly applicable to a pragmatic study of literary texts (Leech, 2008: 118). In the analysis, he demonstrates that the characters systematically violate Grice's maxims and the Cooperative Principle as well. This last principle springs from the first one and consists of the following maxims:

(i) *The TACT Maxim*
Minimize the cost to others, maximize the benefit to others.
(ii) *The GENEROSITY Maxim*
Minimize the benefit to self, maximize the cost to self.
(iii) *The APPROBATION Maxim*
Minimize dispraise to others, maximize praise to others.
(iv) *The MODESTY Maxim*
Minimize praise of self, maximize dispraise of self.
(v) The AGREEMENT Maxim
Minimize disagreement between self and others, maximize agreement between self and others.
vi) The SYMPATHY Maxim
Minimize antipathy between self and others, maximize sympathy between self and others.
Figure 3: Leech, 2008: 92.

Researchers can beneficially use keyword or cluster lists in a pragmatic analysis of a text (e.g. Norgaard et al., 2010: 40). While the focus of corpus linguists is traditionally on an isolated word or groups of words, it is important to look at larger patterns as well (clusters, collocations or whole sentences in the case of the conversational maxims), particularly if they are relevant to its stylistic interpretation. A pragmatic analysis of *AW* where the corpus linguistic approach is used was undertaken for instance by Inaki and Okita (2005) in their article "Alice's roles". They focused on verbs of saying and the collocating adverbials in order to analyse the conversational relationships of Alice and the other characters, her attitude to them, and her status in the *Wonderland* and the *Looking-Glass* stories. The corpus approach points the researcher's attention to significant repetitive and deviant patterns in the text and these can subsequently be pragmatically assessed in a qualitative analysis. A large part of *Alice in Wonderland* is made up by dialogue, as attested by the high frequency of the verbs of saying in the text (see Section 5.4), which leads logically to an inclusion of a pragmatically orientated part at the conclusion of the present analysis. A conversational exploration of the interactions in *AW* can serve to establish the rules of its pragmatic relations as one of the key aspects of the text.

4.2. The data and the hypothesis

The material used for the present analysis is the Project Gutenberg plain text of *Alice's Adventures in Wonderland* (THE MILLENNIUM FULCRUM EDITION 3.0).[22] The text is, of course, cleaned of all additional information, e.g. regarding copyright, etc. The text is 26, 679 words long (27, 189 tokens and 2, 600 types). As a reference to the material, *The Annotated Alice* (Gardner, ed. 2001) is consulted, as it offers valuable information about various references and historical realities of Carroll's world, as well as explanations of his linguistic puzzles.

There are two reference corpora employed, one of them comprising of texts of the Golden Age of children's literature (1860s to 1920s) (see also Section 4.1.1), the other of contemporary Victorian fiction. The fist corpus is necessarily closer in genre to *AW* and can be expected to yield a slightly different keyword list than the second one. The size of the

[22] Posting date: June 25, 2008; release date: March, 1994; last updated: December 20, 2011. Project Gutenberg, May 2015 https://www.gutenberg.org/cache/epub/11/pg11.txt .

children's literature corpus is 390, 494 words (296, 256 tokens and 15, 654 types). The books it includes are the Project Gutenberg editions of:

Kingsley's *The Water-Babies* (1863),
Macdonald's *The Princess and the Goblin* (1872),
Stevenson's *Treasure Island* (1883),
Kipling's *Jungle Book* (1894),
Potter's *The Tale of Peter Rabbit* (1902),
Barrie's *Peter and Wendy* (1911),
Burnett's *The Secret Garden* (1911),
and Milne's *Winnie the Pooh* (1926).

The size of the Victorian fiction corpus is 1, 469, 708 words (1, 266, 036 tokens and 30, 090 types). It includes the Project Gutenberg editions of:

Dickens's *Oliver Twist* (1838)
Emily Brontë's *Wuthering Heights* (1847)
Gaskell's *North and South* (1855)
Collins's *The Woman in White* (1859)
Eliot's *The Mill on the Floss* (1860)
Hardy's *Tess of the D'Urbervilles* (1891)
and Doyle's *The Adventures of Sherlock Holmes* (1892)

As this research is corpus-driven (Tognini-Bonelli, 2001), there is no primary hypothesis. On the basis of other corpus linguistic analyses, however, it can be hypothesized that the keyword analysis will indicate both aboutness and the grammatical structure of the text. Mahlberg (2012) indicated five-word clusters as revelatory character descriptors, and as much can be presumed about the clusters of *AW*. Many of the Wonderland creatures, moreover, are characterized by persistent repetitiveness (see e.g. Sections 5.7.1 and 5.7.8) which, if statistically supported, will serve as a corpus-based description of the text's language. In the pragmatic analysis of selected passages (see Section 5.7), several breaches of Grice's maxims and those of the Cooperative Principle can be expected, owing to the "inconstancy of language" and the "non-responsive", inherently subversive nature of the Wonderland dialogue (Lakoff, 1993: 372, 379).

5. The research

In this section, the results of the analysis are discussed. First, I focus on keywords (Sections 5.1 to 5.6), then on four-word clusters (Sections 5.7.1 to 5.7.8) and lastly on the text's pragmatics (Section 5.8). The keyword lists obtained by the comparison of *Alice in Wonderland (AW)* with the corpus of children's fiction (*ACHcorp*) and contemporary fiction *(ACOcorp)* are displayed in the appendix (Figures 4 and 5), where a list of negative keywords (Figure 6) and clusters (Figure 7) can also be found.

In the keyword sections, the keywords obtained after generation are manually sorted into part of speech categories and further into semantic domains. As this is done manually, it is ruled purely by the researcher's judgment. Where appropriate, I look closer at the concrete concordance lines in order to explore the relationship of the keywords and their context. Character names have not been omitted from the analysis as they can provide valuable information about character depiction in the story. They are the first group of keywords we shall focus on here.[23]

[23] If keywords are meant (as opposed to characters), I use italics. *Alice* therefore means the name as a keyword, whereas Alice is used for the character.

5.1. Lexical items: character names

ACHcorp	**ACOcorp**	**ACHcorp**	**ACOcorp**	**ACHcorp**	**ACOcorp**	**ACHcorp**	**ACOcorp**
A,B,C	**A,B,C**	**D,E**	**D,E**	**F**	**F**	**G**	**G**
Gryphon	Mock Turtle	Hatter	Queen	Alice	Alice	cat(s)	cat(s)
Mock Turtle	Gryphon	King	Hatter	Dinah	Dinah	whiting	whiting
Rabbit	Dormouse	Queen	Duchess			S(s)erpent	S(s)erpent
Dormouse	Rabbit	Duchess	King				lobster(s)
Mouse	Mouse	Knave	Knave				porpoise
March Hare	March Hare	cook	cook				
Caterpillar	Caterpillar	soldiers	Majesty				
Cheshire cat	Cheshire cat		soldiers				
Dodo	Dodo		gardeners				
Pigeon	Pigeon		players				
Lory	puppy						
hedgehog	Lory						
puppy	hedgehog						
Bill	pig						
footman	Lizard						
	Bill						
	guinea-pigs						
	footman						
	flamingo						

Table 1: key names of characters in Alice (for keyness and frequency measures see Figures 4 and 5 in the appendix).

In corpus studies of texts character names necessarily display a high level of keyness since by their unusually high frequency of occurrence (and often their absence from a reference corpus), they are greatly specific. In *AW*, all of 30 characters come out key. Beside characters as such, there are several character references, i.e. names which are referred to in the text, but do not function as full-fledged characters. While the names' keyness is highly influenced by the composition of the reference corpora, their raw frequency in the text tells us how many times they actually appear. Characters like Alice, the Mock Turtle, the Hatter or the Gryphon (to name just a few) have a high level of keyness in both *ACHcorp* and *ACOcorp*, and are very frequent in the text, suggesting that they are infrequent in the reference corpora (0-1 occurrences in the corpus of children's literature) and frequent in the study corpus (397 instances of *Alice* and around 50 instances of *Mock Turtle*, *Hatter* and *Gryphon* each). On the other hand, the Caterpillar and the March Hare's frequencies in the text are not as high as those I have just mentioned, while they are on a very high position in the keyword list. This is because they are both very infrequent in the reference corpora and structurally significant in the study corpus. Less significant characters, who still have a relatively high frequency of occurrence (17) in the study corpus in comparison with other less significant characters, like

e.g. Bill, on the other hand assume a low position on the lists because they are relatively high both in the study corpus and the reference corpora (the corpus of contemporary literature has a frequency of occurrence of the word *Bill* as high as 134). Lastly, there are words like *executioner*, which are low in frequency in both the study corpus (6) and the reference corpora (0 and 1). These are found at the bottom of the lists. These factors have to be taken into consideration when discussing the keyword lists in order not to overrate (or, on the other hand, underrate) the individual words' significance.

Another important point to make is that the keyword list obtained by comparing *AW* to the corpus of Victorian (contemporary) literature generally displays the same items as the *ACHcorp* keyword list, but it also adds some more. This is due to the fact that it is several times bigger. A look at the raw frequencies of names in the study corpus and the reference corpora (see Figures 4 and 5 in the appendix) can show us what names appear in the corpus of children's fiction and that of contemporary literature as well as in *AW*. For instance, the corpus of children's literature displays several instances of the words *Queen, mouse, cat(s), King, rabbit, Bill* or *pigeon*. Frequencies of these words range from 199 of *King* to 7 of *pigeon*. On the other hand, words like *hatter, gryphon* or *Duchess* have no representations. This is given by the fact that people characters like King or Queen are a staple of many fairy tales, as well as animals like cat, mouse or rabbit. The proper name *Bill* can be explained by the fact that it is a common Christian name in the English language. On the other hand, characters like hatters, gryphons or Duchesses can be presumed to be less common. Raw frequencies in the corpus of contemporary literature tell us that words like *King, Queen* or *rabbit* still feature several occurrences, but they are much less common than in the children's literature corpus. Conversely, the frequency of words like *soldiers* or *Bill* has increased (due to the fact that these words are not specific to the fairy tale genre). Proper names like *Dinah* can be expected not to appear in either of the corpora at all (which is indeed the case), since such a name is highly unusual.

All of the character names are revelatory of the study corpus. We know from the numerous occurrences of animal names and characters like King or Queen in the text that we are here concerned with a representative of children's fiction.

The proper name keywords in *AW* can be sorted into several categories, which are discussed in further detail below. These are:

a) common names which serve as proper names
b) common names which serve as proper, premodified by an adjective
c) ordinary common names
d) names of card characters and court functions
e) names of human characters
f) proper names referring to non-Wonderland characters
g) animal references[24]

As opposed to other types of fiction, the majority of lexical items referring to the main characters in *AW* are common names (group A),- specifically different types of animals. They are *made* proper by initial capital letters, and thus serve a dual purpose: that of saying what species an animal is, and providing it with a proper, reference, name. Animal characters are a frequent semantic category in children's fiction and can serve as its important delimitation. As the story takes place in Wonderland, a foreign landscape, the fact that Alice herself has a proper name (Alice) which does not simultaneously define her as a human child[25] becomes highly unusual (although it is not unique). By employing these rules, Carroll questions the role of naming in fictional writing, a theme that is further explored in the *Looking-Glass* story.[26]

Characters with this type of name, which were identified as key in our analysis, are *Lory, Dodo, Pigeon, Lizard, Gryphon, Dormouse, Mouse* and *Caterpillar*. The Lizard is an exception in this category since he is the only one who is also given a second (Christian) name, *Bill*. This could have been done to increase the comicality of Chapter IV, where other characters (the species of which we do not know) are also named (such as *Mary Ann* or *Pat*). Much of this chapter is structured like a comic sketch from a play of which Alice is the unseeing participant. The Christian names could have been employed by Carroll to evoke scenes from the stage where the Rabbit assumes the role of a gentleman and Pat, Mary Ann

[24] See Table 1 above. The categories are marked accordingly as A, B, C, D, E, F, and G.
[25] If Alice were to be named according to the rules of Wonderland naming, she would have to be called the Child or the Girl, and this would be her sole name, providing her with a reference and definition both (e.g. when the Rabbit's name is the Rabbit). However, this is not the case, and there are two possible ways to refer to her in the text. One is "Alice", a term used by Carroll, the other one is "you", "child", "girl" or "young lady", used by the characters.
[26] Gardner (2001: 187) speaks of the wood where things have no names in charter III of *Through the Looking-Glass* as a philosophically insightful construct. By including it in the *Alice* stories, Carroll points to the arbitrariness of naming.

and Bill become his servants.[27] By using this strategy, Carroll in fact humanized the characters involved and made the situation more relatable to his readers.

Excerpt 1: Alice in Wonderland (Gardner, 2001: 42-43.)

Next came an angry voice--the Rabbit's--'**Pat! Pat!** Where are you?' And then a voice she had never heard before, 'Sure then I'm here! Digging for apples, yer honour!'

'Digging for apples, indeed!' said the Rabbit angrily. 'Here! Come and help me out of THIS!' (Sounds of more broken glass.)

'Now tell me, **Pat,** what's that in the window?'

Such names, however, are exceptional. In the *A Mad Tea-Party* chapter, for instance, the Hatter and the March Hare refer to the Dormouse as *Dormouse* even in the vocative and otherwise address each other as *you.*

Excerpt 2: Alice in Wonderland (Gardner, 2001: 77-78.)

'Then the Dormouse shall!' they both cried. 'Wake up, Dormouse!' And they pinched it on both sides at once.

'Two days wrong!' sighed the Hatter. 'I told you butter wouldn't suit the works!' he added looking angrily at the March Hare.

The second category differs slightly from the first one by the fact that the names included in it (*White Rabbit, Cheshire Cat, March Hare, Mock Turtle* and *Frog-Footman/Fish-Footman*) also possess a qualifying premodifier (group B). With the exception of the Mock Turtle, they surface in the text both with and without it. These qualifying adjectives, of course, have a reason to be there. The Cheshire Cat and the March Hare, for instance, have origin in the popular sayings of the day, "grin like a Cheshire cat" or "mad as a March hare" (Gardner, 2001: 62,69). Interesting members of this group are the Fish- and Frog-Footmen, whose names suggest that they subsume common animal name and human occupation in one denomination.

The third category, ordinary common names (group C), is represented by animal characters whose names remain without a capital letter in the text; in the keyword list these include *puppy, guinea-pigs, flamingo, hedgehog* or *pig*. The reason why these characters' names are not capitalized is hard to determine, but it is probably caused both by their fleeting appearance in the text and by the fact that these are not characters per se. Without exception, the common

[27] In *The Annotated Alice* (2001, 39), Martin Gardner says that Mary Ann "was at the time a British euphemism for "servant girl"". The name therefore has a dual function: that of providing reference and a connotation.

name animals do not possess the art of speech and do not assume the level of importance which belongs to the characters of the first and second classes. The puppy is often seen by literary critics like a foreign invader from Alice's world in the landscape of Wonderland (Gardner, 2001: 46) and the flamingo and hedgehog especially are a sort of inferior beings used as instruments in playing croquet. A slight exception might be the guinea-pigs who form a part of the jury and therefore possess a more important role in the story.

In the fourth category, we find card characters (group D), specifically the keywords *King, Queen* (and an umbrella term for both: Your *Majesty*) and *Knave*. There are other card characters as well, however, who are not always denoted by a specific name, but are given umbrella terms: *soldiers*, *gardeners* and *players*. They are all members of the card pack (we know that the gardeners are also called Five, Seven and Two), but in the majority of cases we are not told their number names. It is possible that their identities are exchangeable. The fifth category (group E) is made up by the keywords *Hatter*, *Duchess* and *cook*: the human characters of *AW*. *Hatter* and *Duchess* are again written with a capital letter (once again probably due to their larger role in the text), whereas the *cook* possesses no such privilege. It is also interesting to note that *cook* and *Hatter* denote occupations. *Duchess* had perhaps better be classed in the category of "social position".

Two key characters belong in the sixth category, namely Alice herself and her cat Dinah (group F). These are specific for two reasons. The first is that they are Christian names (uncommon indeed in Wonderland) and that their carriers are strangers in the landscape of the book. *Dinah* has only a minor reference in the text, but *Alice* is more interesting. In the text, she is also referred to by others as "girl" (by the Pigeon, who interestingly has no idea of *what* she might be), "child" by the Queen and the Duchess, and "young lady" by the March Hare, the Mock Turtle and the Hatter. Human children probably exist as an entity in Wonderland in some way or another (the Pigeon talks of seeing "a good many little girls in [his] time) but the identity of the only child we meet in Wonderland changes quickly (it turns into a pig in the *Pig and Pepper* chapter). So, in fact, does Alice's identity, as she constantly grows and shrinks. Both physical and personal identity of children in *AW* seems to be an unstable thing indeed, further supported by the fact that in her conversation with the Caterpillar in chapter V, Alice cannot tell with any certainty who she is (see Excerpt 11 in Section 5.8).

The following word cloud represents the keywords found on the *ACOcorp* list with the word *Alice* at the centre and offers an interesting visualization of the keyword list.

The last category is made up by animal references (group G), i.e. animals who are mentioned in the text but do not surface as actual characters. These include *cat/cats* (sometimes in reference to the Cheshire cat, but also to non-Wonderland cats in the speech of Alice), *whiting*, *S(s)erpent* (used by the Pigeon as a proper name and by Alice as a common name), *Lobster* (in the title of "The Lobster Quadrille") and *lobsters* (the dancers who participate in it), and, interestingly, *porpoise*, who only appears in the text as a malapropism for the word "purpose". This category can be subsumed in some of the earlier ones and need not perhaps be discussed in more detail.

In conclusion of this section, let us recapitulate the various name groups of the text. Firstly, there are common names serving as proper names, such *as Dormouse* or *Caterpillar*. They are written with a capital letter and constitute full-fledged Wonderland characters. The second group is similar to the first one, but animals included in it also possess an adjectival first part of their names (e.g. the *Cheshire Cat* or the *March Hare*). The third category is made up by ordinary common names, such as *puppy* or *guinea-pigs* (without a capital letter) and includes

lesser characters. The fourth category includes names of card characters and court functions, such as *King* or *Queen*. In the fifth category, there are names of human characters like *Hatter* or *cook*, while in the sixth we can find proper names of non-Wonderland characters (*Alice* and *Dinah*). In the last category, there are the names which are only referred to in the text, but whose bearers do not function as characters as such, e.g. *whiting* or *lobsters.* For concrete examples of these categories, see the concordance lines below[28]:

```
he Hatter were having tea at it: a Dormouse was sitting between them, fast
 some time in silence: at last the Caterpillar took the hookah out of its m
u--are you fond--of--of dogs?' The Mouse did not answer, so Alice went on e
n into it: there were a Duck and a Dodo, a Lory and an Eaglet, and sev
w?' she asked the Gryphon, and the Gryphon answered, very nearly in the sam
oor little juror (it was Bill, the Lizard) could not make out at all what h
 was another long passage, and the White Rabbit was still in sight, hurryin
as a little startled by seeing the Cheshire Cat sitting on a bough of a tre
 Come on!'  So they went up to the Mock Turtle, who looked at them with lar
ree in front of the house, and the March Hare and the Hatter were having te
 round the thistle again; then the puppy began a series of short charges at
 and the choking of the suppressed guinea-pigs, filled the air, mixed up wi
ld: but it makes rather a handsome pig, I think.' And she began thinking ov
e.  By the time she had caught the flamingo and brought it back, the fight
nt in search of her hedgehog.  The hedgehog was engaged in a fight with ano
 CHAPTER XI. Who Stole the Tarts?  The King and Queen of Hearts were seated
can you?' he added, turning to the Knave.  The Knave shook his head sadly.
'  'Oh, don't bother ME,' said the Duchess; 'I never could abide figures!'
d she hurried out of the room. The cook threw a frying-pan after her as sh
e could not tell whether they were gardeners, or soldiers, or courtiers, or
eft off quarrelling with the other players, and shouting 'Off with his hea
as she could, 'If you do. I'll set Dinah at you!'  There was a dead silence
 told you that.'  'If I'd been the whiting,' said Alice, whose thoughts wer
phon, 'that they WOULD go with the lobsters to the dance. So they got throw
 wriggling down from the sky! Ugh, Serpent!'  'But I'm NOT a serpent, I tel
a journey, I should say "With what porpoise?"'  'Don't you mean "purpose"?'
```

5.2. Lexical items: nouns

Lexical items, especially nouns, are important indicators of aboutness. If we only glance at the categories a) to f) below, we can see that words related to food, justice, songs, speech or personal possessions appear as significant concerns in the text. The category of personal possessions functions in a similar way as food does: it serves to delineate character specificities. As I shall discuss below, it is important to study the keyword plot tool to see if these expressions are local only and where they appear, or if they are spread out across the whole text. This can help to localize its thematic concerns.

[28] For more on the role of names and characterization in general see keyword Sections 5.2, 5.3, 5.4 and cluster Sections 5.7.2, 5.7.3 and 5.7.8.

The key noun categories in *AW*:

a) food keywords
b) justice keywords
c) song and verse keywords
d) personal possessions keywords
e) speech keywords
f) other keywords

food	personal possessions	justice	song and verse	speech	other
soup	fan	jury	twinkle	tone	MUSHROOM
tarts	gloves	court	soo-oop	voice	CROQUET
pepper	slates	executioner	soup	CONVERSATION	SIZE
bread-and-butter	HOOKAH	JURY-BOX	DANCE	*MORAL*	GAME
		JURORS	CHORUS		SEA
		WITNESS	WOW		RABBIT-HOLE
		TRIAL	OOTIFUL		LESSONS
			WILLIAM		TAIL
					HEAD
					BOTTLE
					THING
					POOL

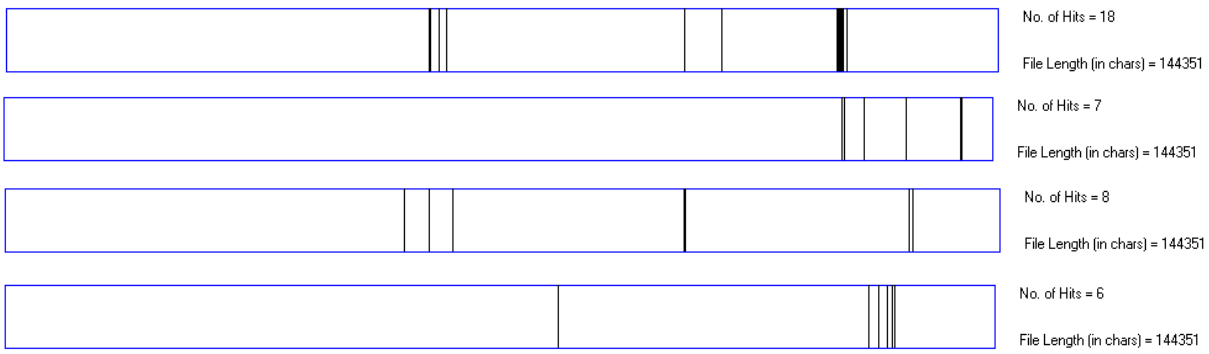
Table 2: key nouns in AW. Words in normal font are common to both the *ACHcorp* and the *ACOcorp* lists, words in capital letters are present only on the *ACOcorp* list and words in bold print and italics are present only on the *ACHcorp* list. (For keyness and frequen measures see Figures 4 and 5 in the appendix).

Let us first discuss the categories of food (see Section 5.4) and personal possessions (see Section 5.5), which both tie in with character depiction. Food includes four items*: soup, pepper, tarts* and *bread-and-butter,* while personal possessions include *gloves, fan, slates* and *hookah*.

The fact that food appears in the text at all could tie in with the fact that it is an important concern for a child audience for whom the book was intended. If we examine the frequencies on the *ACHcorp* and *ACOcorp* lists, we can see that the appearance of all of these items is either marginal or non-existent in the reference corpora (even in the corpus of children's literature). It appears that food just is either not talked about much or that these items especially are specific to *AW* only. However, it needs to be said that the appearance of all of the food items in the text is highly episodic (as evidenced by the keyword plot tool below). *Soup* and *pepper* centre around the *Pig and Pepper* chapter and *soup* appears also later on as a

part of the "Turtle Song" in chapter X. According to Gardner (2001: 62), pepper "suggests the peppery ill temper of the Duchess. *Tarts* are an important item since they form the whole crux of the trial in chapter XI. Their presence in the text might be largely for reasons of rhyme ("Who stole the tarts?" - The Knave of Hearts). As for *bread-and-butter*, a staple ingredient of tea parties, it appears first at the mad tea party and later on the Hatter takes it with him into the courtroom.

The following keyword plots show the occurrence of the words *soup, tarts, pepper* and *bread-and-butter* respectively. Visual representations like this can help us identify isolated and recurrent themes in the text. As we can see, the occurrence of food items in the text is limited to individual chapters.

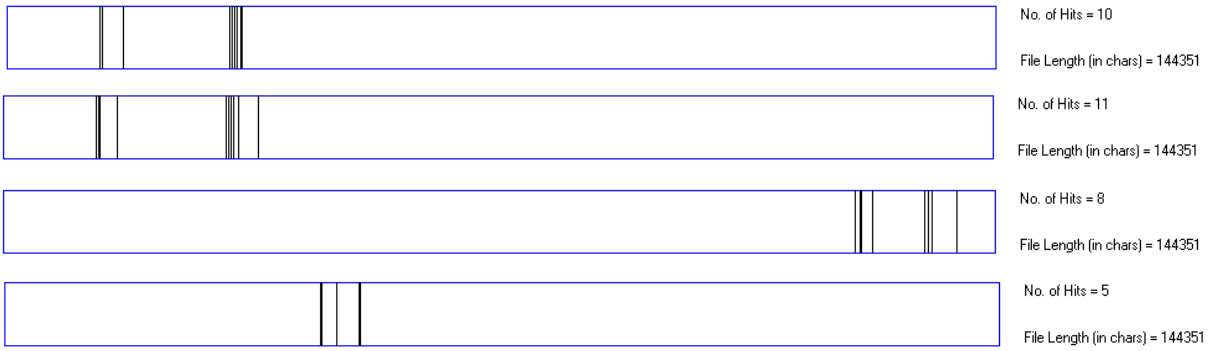


As I mentioned above, food in *AW* serves as an important means of characterization (see also Sections 5.7.2 and 5.7.3) since it is so closely associated with separate episodes and people (the bread-and-butter with the Hatter (incidentally, there is a rhyme there in the same way as there is with the tarts and the Knave of Hearts), the soup with the Mock Turtle, the Duchess and the cook and the pepper with the Duchess and the cook. A similar type of unique characterization can be seen in the personal possessions keywords such as *gloves* and *fan*, which are associated with the White Rabbit, the *slates* connected with the members of the jury and the *hookah*, which is the property of the Caterpillar. These words help to associate the characters with specific objects (be it food or possessions) in the readers' minds and are a means of creating prominence in the text.

she helped herself to some tea and **bread-and-butter**, and then turned to th
one, the cook took the cauldron of **soup** off the fire, and at once set to w
bs, to sing this:-- 'Beautiful **Soup**, so rich and green, Waiting in
voice, 'What are tarts made of?' '**Pepper**, mostly,' said the cook. 'Treacl
he Knave of Hearts, he stole those **tarts**, And took them quite away!'

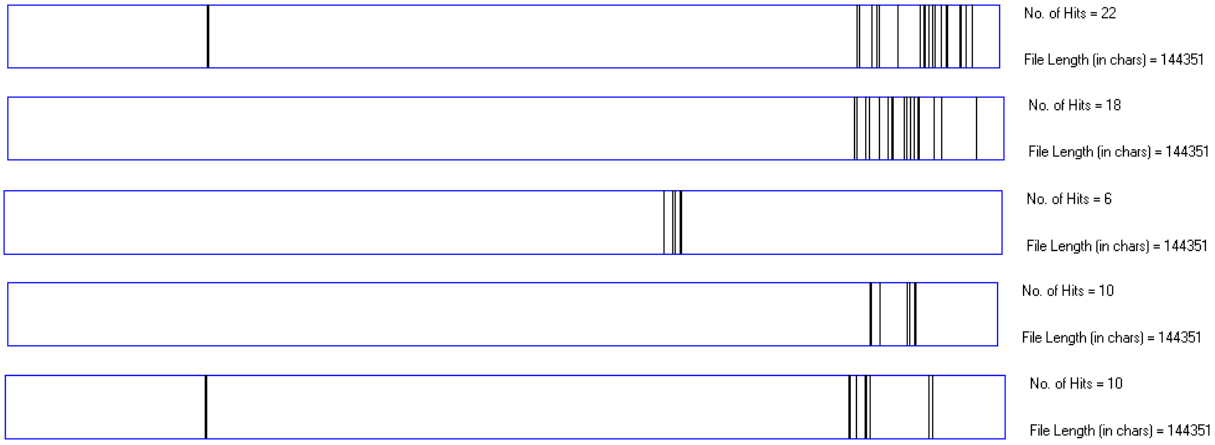
```
dressed, with a pair of white kid gloves in one hand and a large fan in the
  The jury all wrote down on their slates, 'SHE doesn't believe there's an
: at last the Caterpillar took the hookah out of its mouth, and addressed h
```

The keyword plots below again show the distribution of personal possessions in *AW* (*fan, gloves, slates* and *hookah* respectively). The common place of occurrence of *gloves* and *fan* suggest the fact that they co-occur (which is indeed the case, as they are carried together by the Rabbit). We can see that *slates* are limited to the final part of the text (the trial) and that *hookah* occurs at the end of the first third of the text (when Alice meets the Caterpillar).



The discussion of local prominence leads us to another contextually limited group: that of the law of justice, surfacing in the words *jury, court, executioner, witness, trial, jury-box* and *jurors*. As before, these keywords are confined to a specific part of the text only (as can be seen in the keyword plot tool below). The majority surfaces solely in chapter XI, where the Knave of Hearts is tried for theft, except e.g. *jury* and *trial*, which occur also in the Mouse's tale in chapter III and *executioner*, who comes to behead the Cheshire Cat before the commencement of the trial of the Knave of Hearts in chapter VIII. Their high number on the keyword list is indicative of the structural prominence of this chapter in the text. Of course, their presence on the keyword list is also related to their infrequent use (or total absence) from the reference corpora, from which *AW* deviates. The word *court*, for instance, occurs 69 times in *AW* as opposed to 18 times in the corpus of contemporary fiction. The comparison of *AW* and the corpus of children's literature also displays some interesting statistics: the word *jury* occurs only once there as opposed to 17 times in *AW*, *witness* and *trial* have the frequency of 3 in the corpus of children's literature versus 10 in *AW*. The high frequency of words relating to justice points it out as an important indicator of aboutness in the text and signals out this chapter as a pivotal turning point of the story.

The keyword plot of the words *jury*, *court*, *executioner*, *witness* and *trial* (respectively) shows that all of these occur predominantly in the latter part of the text. Their occurrence is therefore limited only.



The song and verse keywords (see Section 5.7.7) indicate that singing and recitation is an important concern in Wonderland and that it occurs frequently. They form a substantial part of the text (hardly a chapter goes by without a song or a rhyme) and become key not only because some of them are not likely to appear in other texts as well, but also because of the inherent nature of songs and verses: repetition. Even a short song or verse can become key in the text if it features large-scale repetition. Leech (2008) calls this effect "the sequential exploration of sameness". In the songs of *AW*, the repetition of nonsensical rhythmical patters creates humorous nonsense poetry, such as in the following mock-pathetic excerpt from the "Turtle Song" (Excerpt 3). The dashes dividing individual words in the latter part of the song indicate both the manner in which the original song ("Star of the Evening") was performed and the intermittent sobs of the Mock Turtle (Gardner, 2001: 112). The excerpt on the right (Excerpt 4) is taken from the Hatter's "Twinkle, twinkle, little bat" from chapter VII.

Excerpts 3 and 4: Alice in Wonderland (Gardner, 2001: 112 and 76-77.)

*'**Beautiful Soup**, so rich and green,*
Waiting in a hot tureen!
Who for such dainties would not stoop?
Soup of the evening, beautiful Soup!
Soup of the evening, beautiful Soup!
Beau--ootiful Soo--oop!
Beau--ootiful Soo--oop!
Soo--oop of the e--e--evening,
Beautiful, beautiful Soup!

"Twinkle, twinkle, little bat!
How I wonder what you're at!"
"Up above the world you fly,
Like a tea-tray in the sky.
Twinkle, twinkle--"'
Here the Dormouse shook itself, and began singing in its sleep **'Twinkle, twinkle, twinkle, twinkle--'**

and went on so long that they had to pinch it to make it stop.

The keyword *dance* is repeated in the "Lobster Quadrille" song, while *soup* (and its equivalent *soo-oop*) and the word *ootiful* (standing for the latter part of be*autiful*) are frequent in the "Turtle Soup" song in the same chapter. *Chorus* appears in several songs but clusters especially in the Duchess's lullaby (together with the word *wow*). *William* achieves prominence for two reasons. Firstly, it is repeated in the "You Are Old, Father William" poem the Caterpillar asks Alice to repeat, secondly, it is repeated several times in the Mouse's lecture on William the Conqueror. As Gardner (2001) says, Carroll's songs are parodies on current popular productions, and they must have provided the readers who knew their real-life counterparts with much amusement. Needless to say, they are important carriers of nonsense. Songs and verses occur in various chapters of the text, where they are relatively evenly distributed. Their appearance in individual songs and verses only determines their isolation.

The keywords mentioned so far have shown us that food and possessions are a means of character association in the text, the justice keywords point to a prominent chapter and song keywords surface largely due to their repetition. If we move on to the next semantic noun group, however, we are confronted with yet another means by which words can become key: pervasive stylistic preference.

Keywords *tone* and *voice* are a frequent feature of Carroll's style of writing, and are indicative of his stylistic preferences (see Sections 5.4 and 5.7.2, 5.7.3). *Tone* is used in the basic pattern '*verb of saying- interlocutor- in- a(n)- *- (*)- tone*.'[29] The asterisks stand for adjectives, some of which are mentioned in the concordance lines below. Beside them, other adjectives found in this construction are *sorrowful, offended, angry, subdued, encouraging, very humble, low* and *hurried, frightened, rather complaining, very hopeful, pleased, deep* and *hollow, very decided, impatient* or *very respectful*. The adjective *offended* rises to the greatest frequency: it is used in the text four times (and also appears key in the keyword list).

```
'  'Well, perhaps not,' said Alice in a soothing tone: 'don't be angry abou
oke.  'As wet as ever,' said Alice in a melancholy tone: 'it doesn't seem t
. 'Poor little thing!' said Alice, in a coaxing tone, and she tried hard to
ot used to it!' pleaded poor Alice in a piteous tone. And she thought of he
ll, be off, then!' said the Pigeon in a sulky tone, as it settled down agai
 handed over to the other, saying, in a solemn tone, 'For the Duchess. An i
```

[29] For information about uninterrupted patterns with variable members see e.g. Starcke, 2010: 117-118.

m I to get in?' asked Alice again, in a **louder** tone. 'ARE you to get in at
er since that,' the Hatter went on in a **mournful** tone, 'he won't do a thing
  'I couldn't help it,' said Five, in a **sulky** tone; 'Seven jogged my elbow.
the boots and shoes!' she repeated in a **wondering** tone. 'Why, what are YOU

The other basic pattern for the word *tone* is '*verb of saying- interlocutor- in- a – tone – of * (*)'.* (The asterisks again stand for adjectives). Specific concordances are given below. As we can see, they are rather more dramatic: adjectives such as *great* and *deepest* are used as noun intensifiers.

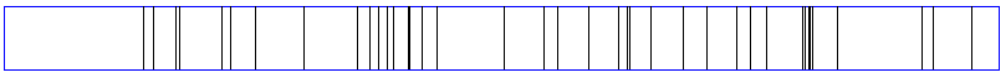
y head's free at last!' said Alice in a tone of **delight**, which changed into
ely story indeed!' said the Pigeon in a tone of **the deepest contempt**. 'I've
hat are they made of?' Alice asked in a tone of **great curiosity**. 'Soles an
  'Wouldn't it really?' said Alice in a tone of **great surprise**. 'Of course
 I BEG your pardon!' she exclaimed in a tone of **great dismay**, and began pic

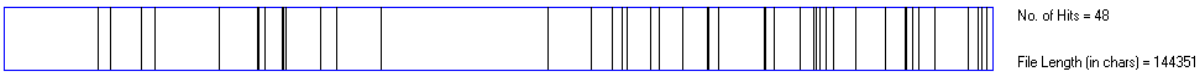
The keyword *voice* is used in a very similar way in the text, the most common pattern being once again '*verb of saying- interlocutor- in- a(n)- *- (*)- voice'.* Some of the adjectives found in this pattern are *low*, *trembling*, *sleepy*, *melancholy*, *feeble*, *loud*, *weak* or *indignant.*

like the Queen?' said the Cat in a **low** voice. 'Not at all,' said Alice: 's
and a Canary called out in a **trembling** voice to its children, 'Come away, m
and addressed her in a languid, **sleepy** voice. 'Who are YOU?' said the Cate
t!' Alice replied in a very **melancholy** voice. 'Repeat, "YOU ARE OLD, FATHE
t asleep,' he said in a hoarse, **feeble** voice: 'I heard every word you fello
ngs!' Alice began in a **loud, indignant** voice, but she stopped hastily, for
sty!' the Duchess began in a low, **weak** voice. 'Now, I give you fair warnin

We can see that Carroll is fond of qualifying his verbs of saying by an accompanying emotive appendix. This is probably given by the fact that it is the only way of identifying the tenor of an utterance in the written text. It serves to dramatize the dialogue in a very simple and accessible way and it is therefore only natural that it should appear so frequently in a work of children's literature. We shall see a similar tendency displayed by the adjectives and adverbs following the verbs of speaking in Section 5.3. For further discussion of this theme, see Sections 5.4. and 5.7.3.

As the keyword plots below show, *tone* and *voice* are not isolated, but evenly spread out across the text. This is given by their nature as frequent modifiers of verbs of saying.





The last keyword I shall mention here is *size*, which serves mainly as an indicator of aboutness. In *AW*, the heroine changes her size several times[30], perhaps echoing the changefulness of childhood and the instability of physical and personal identity.[31] The following keyword plot can help us identify where exactly in the text it happens, and how frequently.

5.3. Lexical items: adjectives and adverbs

ACHcorp adverbs	ACHcorp adjectives	ACOcorp adverbs	ACOcorp adjectives
timidly	large	very	little
anxiously	mad	anxiously	curious
hastily	offended	timidly	large
very		quite	offended
		hastily	mad
		again	executed

Table 3: adjectives and adverbs in Alice (for keyness and frequency measures see Figures 4 and 5 in the appendix).

I have already mentioned Inaki and Okita's (2005: 283) article about the varying roles of the heroine in the two *Alice* books, where Alice remains in "the passive state in *Wonderland*" and turns into "an active explorer in *Looking-Glass*." Talk exchanges of Alice and other characters (see Section 5.8) form the backbone of their article, although the authors also look at the modifiers tied with the node *Alice* to determine how Carroll characterizes her. They found that Alice is frequently associated with such adverbs such as "anxiously, desperately, humbly, meekly, quietly, sadly, and shyly" in the *Wonderland* story, which together with the frequent pre-modifier *poor* points to her submissive status in the first book (Inaki & Okita, 2005: 290, 292).

[30] Gardner (2001: 17) says that Alice changes size all of twelve times in the text.
[31] See Leach & Hollingsworth, 2009: 34.

Three adverbs surfaced as key on both the *ACHcorp* and *ACOcorp* keyword lists in the present analysis, and these were *anxiously, timidly,* and *hastily.*[32] While *anxiously* is certainly used in connection with Alice, it is also frequent with the Rabbit (a predominantly anxious animal) or the King (where it serves to establish a contrast between him and the Queen). As the concordances below show, *anxiously* occurs predominantly with verbs of looking, such as *keep [her] eyes fixed on, peer about, glance, look* or *peep.* In contrast, it is used only twice with verbs of saying.

!' She ate a little bit, and said anxiously to herself, 'Which way? Which
in the middle. Alice **kept her eyes** anxiously **fixed** on it, for she felt sure
o make herself useful, and **looking** anxiously about her. 'Oh, do let me help
ing slowly back again, and **looking** anxiously about as it went, as if it had
t; and while she was **peering about** anxiously among the trees, a little shar
f her head!' Alice **glanced** rather anxiously at the cook, to see if she mea
unted again, and Alice **looked** very anxiously into its face to see what was
 At this moment Five, who had been anxiously **looking** across the garden, cal
 the White Rabbit, who was **peeping** anxiously into her face. 'Very,' said A
 in a low, hurried tone. He **looked** anxiously over his shoulder as he spoke,
s?' asked the Mock Turtle a little anxiously. 'Yes,' said Alice, 'we lear
g put on his spectacles and **looked** anxiously round, to make out who was ta
aid--' the Hatter went on, **looking** anxiously round to see if he would deny
,' said the cook. The King **looked** anxiously at the White Rabbit, who said

The adverb *timidly* occurs overwhelmingly with the node *Alice*, mostly in the context where the Cheshire cat appears, suggesting that this character has a rather intimidating position in the text. As well as *anxiously, timidly* occurs in the text in connection with the King. It occurs most frequently with the verb *said*, though *began* and *went/walked up* to are also used. *Hastily* occurs ten times out of sixteen instances with *Alice*, suggesting that it is again one of the characteristic expressions connected with her status in the text. If we examine the concordances in more detail, we can see that *hastily* describes her usual behaviour following a blunder. This is especially clear in connection with *checked herself* and *stopped*, but it surfaces overall.

tering of feet in the distance, and she hastily dried her eyes to see what
fan she was holding, and she dropped it hastily, just in time to avoid shri
t. 'Oh, I beg your pardon!' cried Alice hastily, afraid that she had hurt t
to save her neck from being broken. She hastily put down the bottle, saying
, I'm not particular as to size,' Alice hastily replied; 'only one doesn't
 matters a good deal to ME,' said Alice hastily; 'but I'm not looking for e
 the March Hare went on. 'I do,' Alice hastily replied; 'at least--at leas
y 'I once tasted--' but **checked herself** hastily, and said 'No, never') '--s
een them at dinn--' she **checked herself** hastily. 'I don't know where Dinn

[32] Other collocating adverbs of the node *Alice* include *politely, certainly, eagerly, thoughtfully, indignantly, cautiously, decidedly, loudly, angrily, hardly, sharply, humbly, dreadfully, doubtfully* and *suddenly.*

loud, indignant voice, but she **stopped** hastily, for the White Rabbit cried

The fact that together with *politely, thoughtfully, cautiously, humbly* and *doubtfully*, submissive adverbs predominate over the more assertive ones, such as *angrily*, *sharply, indignantly* or *decidedly* (none of which surface as key) supports Inaki and Okita's (2005) view of Alice as a largely submissive heroine in *Wonderland* with, let us add, occasional spurts of assertiveness. The keyness of these adverbs also points to Carroll's fondness for qualifying verbs with an appropriate adverb, a stylistic feature which is instantly noticeable in the text.

A short note suffices on *very*, the last adverb found on both the keyword lists. It premodifies either adjectives or other adverbs and its keyness in the text together with a relatively high frequency (144 hits) probably stems from the fact that it is a basic and very common intensifier of these word classes. Frequency alone (more than twice higher than in the corpus of contemporary fiction and the corpus of children's fiction, too) is enough to justify its keyness, although the following excerpt, which shows instances of *very* within 235 words of text (the excerpt given here is abbreviated) displays rather too plentiful occurrences of it.

Excerpt 5: Alice in Wonderland (Gardner, 2001: 11-12.)

There was nothing so **VERY** remarkable in that; nor did Alice think it so **VERY** much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!'

(…)

The rabbit-hole went straight on like a tunnel for some way, and then dipped suddenly down, so suddenly that Alice had not a moment to think about stopping herself before she found herself falling down a **very** deep well. Either the well was **very** deep, or she fell **very** slowly, for she had plenty of time as she went down to look about her and to wonder what was going to happen next.

Let us now turn to the exploration of the text's adjectives. Two semantically connected adjectives existing on opposite sides of a scale in the keyword lists are *large* and *little*, both of which echo the size motif in the text (see Section 5.2) and prove it to be a structurally important element. *Large* seems to be favoured by Carroll, since oddly enough, *big* does not occur in *AW* at all. It is interesting to mention that the frequency of *little* (128) far predominates over *large* (33), probably because it is not used solely as an indicator of size, but also as a quantifier (e.g. in sentences like "She was a little nervous" or "said Alice, a little timidly"). The collocates of *large* include *eyes, saucepan, rose, ring, rabbit-hole* or *pool*,

while the collocates of *little*, which are more frequent and numerous, include *thing, door, golden key, way, timidly, pattering, house* or *girls*.[33] The adjective *offended* ties in with the keyword *tone* (mentioned in Section 5.2). As the following concordance lines show, a majority of the occurrences of *offended* do indeed happen with the word *tone*. Other collocations of *offended* include modals (*must be* offended, *would be* offended) and the adverbial phrase *so easily* offended. This adjective points to the centrality of conflict in the text, where purposely-impolite things are said in order to challenge existing linguistic rules (see Section 5.7).

she felt certain it **must be really offended**. 'We won't talk about her any m
a sorrowful tone, 'I'm afraid **I've offended** it again!' For the Mouse was sw
going to say,' said the Dodo **in an offended tone**, 'was, that the best thing
sper, half afraid that it **would be offended** again. 'Mine is a long and a s
poor Alice. 'But you're **so easily offended**, you know!' The Mouse only gro
he creatures wouldn't be **so easily offended**!' 'You'll get used to it in ti
nothing yet,' Alice replied **in an offended tone**, 'so I can't take more.'
ay,' the Mock Turtle replied **in an offended tone**. And the Gryphon added 'Co
that the Gryphon said, **in a rather offended tone**, 'Hm! No accounting for ta
'It's a pun!' the King added **in an offended tone**, and everybody laughed, 'L

Mad and *curious* revealingly tell us something about the founding elements of Wonderland. There, animals and people are all mad (as the Cheshire cat says)[34] or can go mad at any moment, just like the March Hare. *Mad* indeed does cluster around the first meeting of Alice with the Cheshire cat, where this word is repeated several times in a short space of time.

Excerpt 6: Alice in Wonderland (Gardner: 2001, 67-68.)

'In THAT direction,' the Cat said, waving its right paw round, 'lives a Hatter: and in THAT direction,' waving the other paw, 'lives a March Hare. Visit either you like: they're both **mad**.'

'But I don't want to go among **mad** people,' Alice remarked.

'Oh, you can't help that,' said the Cat: 'we're all **mad** here. I'm **mad.** You're **mad.**'

'How do you know I'm **mad?**' said Alice.

'You must be,' said the Cat, 'or you wouldn't have come here.'

Alice didn't think that proved it at all; however, she went on 'And how do you know that you're **mad?**'

'To begin with,' said the Cat, 'a dog's not **mad.** You grant that?'

'I suppose so,' said Alice.

'Well, then,' the Cat went on, 'you see, a dog growls when it's angry, and wags its tail when it's pleased. Now I growl when I'm pleased, and wag my tail when I'm angry. Therefore I'm **mad.**'

[33] All of these occur within the first 10 most frequent collocates of *large* and *little*.
[34] Gardner (2001: 68).

Curious mostly describes Alice's feelings about several situations that happen when she falls through the rabbit-hole. Its left-hand collocates frequently include verbs of the senses, such as *feeling, feel, see* and *notice.* Keyword *executed* has six hits in the concordances, half of them belonging to the Queen, half of them to the King, who repeatedly says "…or I'll have you executed." This lends the text a dark atmosphere, which Kincaid (1973: 92) mentioned in his essay on "Alice's Invasion of Wonderland", and jars strangely with the adverbs normally found with the King: *timidly* or *anxiously*. An explanation for this could be that the King in fact has two distinct personalities in the text: a timid one, which he displays before the trial (and very probably in the vicinity of his spouse), and one he assumes for the purpose of the trial: that of a stern, merciless judge.

ole--and yet--and yet--it's rather **curious**, you know, this sort of life! I
s!' she thought. 'But everything's **curious** today. I think I may as well go
hem red. Alice thought this a very **curious** thing, and she went nearer to wa
long breath, and said 'That's very **curious**.' 'It's all about as **curious** as
don't be nervous, or I'll have you **executed** on the spot.' This did not see
epeated angrily, 'or I'll have you **executed**, whether you're nervous or not.
marked the King, 'or I'll have you **executed**.' The miserable Hatter dropped

5.4. Lexical keywords: verbs

ACHcorp lexical verbs	ACOcorp lexical verbs	ACHcorp grammatical verbs	ACOcorp grammatical verbs
Said	said	m (standing for am)	m (am)
Replied	began	t (negative auxiliaries and modals)	re (are)
remarked	went		t (negative auxiliaries and modals)
added	eat		doesn (standing for doesn't)
	sneezing		
	thought		
	added		
	know		
	getting		

Table 4: lexical and grammatical verbs in Alice (for keyness and frequency measures see Figures 4 and 5 in the appendix).

Verbs of saying form a distinctive group in *AW*, including the verb *said, replied, remarked, added,* and, to a partial degree, *went (on)* and *began (talking, to say/repeat,* etc.). These verbs can be expected to occur with great frequency in other works of literature as well, but in *AW* they are computed as key, suggesting their centrality in the narrative. A definite leader between them in terms of frequency is *said*, with 462 occurrences. This is because it is "the most prototypical reporting verb" (Semino & Short, 2014: 36). How frequent *said* is is

especially evident in the *ACHcorp* keyword list. Although this reference corpus is approximately 15 times bigger, it contains only 7 times more occurrences of the word *said*. This means that *AW* employs the word about twice more frequently. *Replied* appears in the text 29 times, *added* 23 times and *remarked* 10 times, which, compared to the frequencies in the corpus of children's literature, are also high numbers (considering the sizes of the study corpus and the reference corpus).

I have already touched upon some of the prototypical patterns of verbs of saying in Sections 5.2 and 5.3, for instance the tendency to be followed by the prepositional phrase *in a * tone/voice* or an adverb of manner. An interesting collocation of various verbs of saying (and thinking) is *to herself* (42 occurrences), which points to the role of introspection in the text (see Section 5.7.2). This introspection is a specific feature of Alice and is not shared with any significant frequency by the other characters.[35] Semino and Short (2014: 118) label this type of introspectiveness as "(free) direct thought". As we can see in some of the concordances below, it is not always clear whether the speech is articulated out loud or not (e.g. "[Alice] went on saying to herself…"), but Semino and Short (2014: 118) group these cases under "(free) indirect thought" anyway because "it is conceivable that the characters have mentally articulated their thoughts in verbal form".[36]

```
ll past it.  'Well!' thought Alice to herself, 'after such a fall as this,
 rather sleepy, and went on saying to herself, in a dreamy sort of way, 'Do
t might end, you know,' said Alice to herself, 'in my going out altogether,
e in crying like that!' said Alice to herself, rather sharply; 'I advise yo
e a little bit, and said anxiously to herself, 'Which way? Which way?', hol
istmas.'  And she went on planning to herself how she would manage it. 'The
 can go back by railway,' she said to herself. (Alice had been to the seasi
 quiet thing,' Alice went on, half to herself, as she swam lazily about in
```

The verb *went* appears in 83 concordance hits, out of which about half are instances of *went on*. Although a majority of these are indeed used in the sense "continued to speak", it is also used in the sense "continued to do something" in general, such as *went on growing* or *went on, taking first one side and then the other*. As the other verbs of saying, it is frequently qualified by an adverb or a prepositional phrase. The verb *began* is used in the text with a direct object (which can be preceded by the prepositions *by* or *in*). Verbs of saying following

[35] The sole exception is the White Rabbit, who mutters to itself and says things to itself in its constant agitation to obey the commands of the Queen and the Duchess. Other characters, with whom introspection occurs only once, are the Dormouse, the March Hare and the Gryphon.

[36] The authors also mention Cohn (1978: 58) who called this speech act "quoted monologue" and considered it the earliest kind of use of "(F)DT" ((free direct thought)) (Semino & Short, 2014: 119).

began are either gerundial (*talking, telling*) or infinitival (*to say, to repeat*). One possible explanation for the presence of verbs such as *went on* or *began* in the text is the constant presence of various forms of interruption (be it of action or dialogue) in *AW*, after which the interrupted utterance has to be either continued or begun anew. Both are connected with uncertainty in speech, displayed by various characters in a stressful or puzzling situation. *Went on* also appears in situations when one or the other participant of the conversation refuses to listen. This fact points again to the uncooperativeness of Wonderland dialogue (see Section 5.7).

id nothing, but looked at Two. Two began **in a low voice**, 'Why the fact is,
ne day, your Majesty!' the Duchess began **in a low, weak voice**. 'Now, I giv
oor man, your Majesty,' the Hatter began, **in a trembling voice**, '--and I ha
nothing else to do, so Alice soon began **talking again**. 'Dinah'll miss me v
introduced to a lobster--' (Alice began **to say** 'I once tasted--' but check
rand words to say.) Presently she began **again**. 'I wonder if I shall fall r
ago anything had happened.) So she began **again**: 'Ou est ma chatte?' which w
and I've tried hedges,' the Pigeon **went on, without attending to her**; 'but
nse in your knocking,' the Footman **went on without attending to her**, 'if we
t to her, so **she took courage, and went on again**:-- 'I didn't know that Ch
seemed not to be listening, so **she went on again**: 'Twenty-four hours, I THI
ut it **puzzled her too much, so she went on**: 'But why did they live at the b
e it would all come wrong, and **she went on in a trembling voice**:-- 'I pa

Thought is used mostly in free indirect representation of Alice's thoughts, though there are a very few instances of phrasal verbs such as *think it over*. The preposition *about* is also used a few times in the text and the reflexive use (*thought to herself*) is present all of 5 times out of 74 concordance hits. An interesting point to make is that this verb is also used several times with an emotionally coloured term *poor Alice* or *the poor child* as opposed to the neutral term Alice. If we look closer at the concordance plot of *poor Alice*, we can see that it is used almost exclusively in the first third of the text, after which there are only two more hits in the latter thirds. Whereas the voice of the narrator and his judgment are present at the beginning of the text, they quickly disappear as the adventure continues.

dear! I shall be late!' (when she **thought it over** afterwards, it occurred t
rds as she fell past it. 'Well!' **thought Alice to herself**, 'after such a f
dead silence instantly, and Alice **thought to herself**, 'I wonder what they W
of WHAT? The other side of WHAT?' **thought Alice to herself**. 'Of the mushro
ven if my head would go through,' **thought poor Alice**, 'it would be of very
wo people. 'But it's no use now,' **thought poor Alice**, 'to pretend to be two
'and things are worse than ever,' **thought the poor child**, 'for I never was
'It was much pleasanter at home,' **thought poor Alice**, 'when one wasn't alwa

The verb *know* occurs in the text in 87 concordance hits, out of which 10 are instances of *don't know*. It is therefore possible to say that it appears mostly in positive statements. *Doesn't know* does not occur in the text at all. A noticeable thing is, however, that a third of all the hits of *know* are comment clauses *you know*, traditionally designed to build rapport between two interactants in a spoken conversation. The comment clauses point clearly to the importance of spoken dialogue in *AW*.

```
ld all have our heads cut off, you know. So you see, Miss, we're doing our
But you're so easily offended, you know!'  The Mouse only growled in reply
oesn't like changing so often, you know.'  'I DON'T know,' said the Caterp
it,' said Alice.  'It goes on, you know,' the Hatter continued, 'in this w
nock, and I could let you out, you know.' He was looking up into the sky a
be what he did with the tarts, you know--'  'But, it goes on "THEY ALL RET
 A little bright-eyed terrier, you know, with oh, such long curly brown ha
'They couldn't have done that, you know,' Alice gently remarked; 'they'd h
that saves a world of trouble, you know, as we needn't try to find any. An
 to nobody, which isn't usual, you know.'  'Who is it directed to?' said o
```

The theme of dialogue is also connected with the grammatical verbs found key in the text, which include the contracted forms *'m* standing for *am* or *'t* standing for *doesn't*. These point to the frequent presence of dialogical contractions in the places where speakers would have naturally made them in spoken conversation. The following excerpt is an example of the result of searching the text for the sign ', which computes all instances of ' followed by a space. This enables us to give an approximation of the number of instances someone speaks in the text. The concordance plot of ' below presents a visualization of how frequently it is employed.[37]

Excerpt 7: Alice in Wonderland (Gardner, 2001: 94-95.)

I can't tell you just now what the moral of that is, but I shall remember it in a bit.'

'Perhaps it hasn't one,' Alice ventured to remark.

'Tut, tut, child!' said the Duchess. 'Everything's got a moral, if only you can find it.'

[37] As AntConc software (Anthony, 2014) displayed more reliable results, it was used in the present computation in preference to Wordsmith (Scott, 2015), where search of ' included several instances of the apostrophe as well.

The verb *eat* (see also Section 5.2 for "food" nouns) appears 18 times in the text, exclusively in the first half of it, after which eating is not discussed any more. Its frequency is increased by the repetitive cluster "Do cats eat bats?" which is a question Alice asks herself several times while falling through the rabbit-hole. Another reason for its appearance in the text are the cakes which enable Alice to change in size. The last time eating is mentioned is the mad tea party, where eating and drinking happens in an endless loop of perpetually arrested time. Collocations of the verb *eat* in *AW* include *bats, birds, cats, eggs, it, her, me, cakes, anything, something, comfits* or *some of the other bit*. An unexpected element indeed is the inclusion of personal pronouns like *her* or *me,* which form a linguistic joke and would be unusual indeed to find in most other works of literature.

The conclusion to be made about the key verbs in *AW* is that verbs of saying clearly dominate, followed by verbs of cognition (*know, think*) and other verbs specific to the nature of the text (*eat*). Verbs such as *say* or *think* are used frequently with a reflexive pronoun "to herself", pointing to the centrality of self-reflection of the heroine in the narrative. The verbs *go on* or *begin*, on the other hand, point to dialogue commencements, interruptions or dialogical uncooperativeness. The verbs of saying and the frequent use of the comment clause *you know* accentuate the fact that there is a large amount of dialogue of *AW*, which is also supported by the frequent presence of contractions. The verbs of saying are frequently followed by a qualifying postmodifier, which describes the tenor of the conversation and helps to dramatize it in a simple and effective way.

5.5. Grammatical items: pronouns

pronouns in AW
she
it
herself
its

Table 5: pronouns in Alice (for keyness and frequency measures see figures 4 and 5 in the appendix).

The pronouns in *AW* are largely predictable. Since the story has a narrator, Alice is referred to in the third person. The key pronouns in the text are therefore *she* and *herself*.[38] Other

[38] Their frequency is further increased by reference to the Queen, who is likewise structurally prominent in the text.

prominent pronouns include *you, it* and *its. You* assumes prominence as a deictic pronoun in speaker interactions. However, it is likewise used as a contact dative in phrases such as *you see, you know* (see Section 5.4). *You know* is the most frequent collocate of *you*, used in the text 45 times, while *you see* is slightly less common (used in the text 11 times).

Keyword *it* assumes various functions in the text, such as anaphoric, cataphoric or deictic reference, as well as an empty subject of a sentence, etc.[39] These functions of *it* are a general property of English and can be therefore presumed to exist in a similar degree in the texts of our referential corpora. The reason for why *it* appears on the keyword list, however, probably springs from the fact that it is also used in reference to Wonderland characters (its frequency in *AW* is 2,19 as opposed to 1,44 in children's literature corpus and 1,28 in the corpus of contemporary Victorian fiction).[40] Its use in reference to the animals is inconsistent, as e.g. in the Caucus-Race chapter, where every animal is referred to as *it*, while the Frog-Footman in chapter VI is referred to consistently as *he*. This may be so because his position as a footman predominates over his species in the interaction with Alice. In the cases of most other animals, the usage of *it(self)* and *he(imself)* (and, of course, *its* and *his*) is mixed, as in the following concordance lines (*its* paws but *his* turn used in reference to the Gryphon, *its* head but *his* eyes used in relation to the Dormouse, etc.). We can see that the use of *it* in the case of the Wonderland animals is by no means straightforward and not easily explained, and would require further in depth analysis. The use of *it* as an empty subject necessarily increases its frequency, but it is still noteworthy that *it* and *its* surface among the highest-ranking keywords in the text, while pronouns such as *he, his* and *him* appear on the list of negative keywords (see Figure 6 in the appendix).

ice ventured to say. 'What is it?' The Gryphon lifted up both **its paws** in
say.' 'So he did, so he did,' said the Gryphon, sighing in **his turn**; and b
The March Hare took the watch and looked at it gloomily: then **he** dipped it
ner!' ('I only wish it was,' the March Hare said to **itself** in a whisper.)
he [the Rabbit] came trotting along in a great hurry, muttering to **himself**
out of the way to hear the Rabbit say to **itself**, 'Oh dear! Oh dear! I shall
ed a little hot tea upon its nose. The Dormouse shook **its head** impatiently
pinched it on both sides at once. The Dormouse slowly opened **his eyes**. 'I

As we can see below in Table 6, collocates of the pronoun *its* comprise of various semantic groups, the most prominent one being words referring to body parts. This is due to the fact that English uses possessive pronouns as their determiner. As the text in fairly short, many of

[39] See e.g. Dušková, 2012: section 4.15 at http://emsa.ff.cuni.cz/4.15.
[40] This is a frequency in decimal numbers (see the % and RC% columns in Figures 4 and 5 in the appendix).

the body collocates appear only once, but they still help to explain the keyness of the word *its*. A tendency which is clearly visible in the body collocates is a focus on the face, where mouth, head and eyes assume prominence. However, the frequency of *its* mouth is greatly augmented by the fact that "[put, took] the hookah out of its mouth" is a repetitive cluster of the *Advice from a Caterpillar* chapter. Its importance in the text should therefore not be unduly exaggerated, although its role in the characterization of the Caterpillar is undeniable (see Section 5.2).

Body collocates	Frequency	Other collocates	Frequency
mouth	6	voice	2
head	6	waistcoat	1
eyes	6	undoing	1
feet	3	sleep	1
face	3	share	1
arms	3	nest	1
tail	2	meaning	1
Right (ear, paw)	2	little	1
wings	1	hurry	1
tongue	1	great	1
paws	1	full	1
nose	1	dinner	1
neck	1	children	1
legs	1	axis	1
forehead	1	age	1
eyelids	1		
ears	1		
body	1		

Table 6: collocates of its in Alice

We can see that Carroll favours the use of the pronoun *it* (and its variants) to the pronoun *he* (and its variants). This may be because however humanized his characters are, they are still seen as animals, and this is a traditional way of referring to them.[41] The keyness of the pronouns *she* and *herself* is predictable, since the use of them in reference to the heroine is frequent in the text.

[41] An interesting point to note is that the situation would be different in numerous translations of A*W* into other languages, such as Czech, where animals would be referred to as *he* or *she* according to whether the nouns are assumed masculine or feminine.

5.6. Grammatical keywords: prepositions, adverbs and articles

ACHcorp	ACOcorp
off	off
the	the
	down
	about

Table 7: prepositions, adverbs and articles in Alice (for keyness and frequency measures see Figures 4 and 5 in the appendix).

Keyword *the* appears on both keyword lists, which suggests that its presence there is not mere chance. That such a high-frequency item in general language should be computed key is unusual, but an excerpt from chapter XI below might explain why this is the case:

Excerpt 8: Alice in Wonderland (Gardner, 2001: 118-119.)

'I'm a poor man, your Majesty,' **the** Hatter began, in a trembling voice, '--and I hadn't begun my tea--not above a week or so--and what with **the** bread-and-butter getting so thin--and **the** twinkling of **the** tea--'

'**The** twinkling of **the** what?' said **the** King.

'It began with **the** tea,' **the** Hatter replied.

'Of course twinkling begins with a T!' said **the** King sharply. 'Do you take me for a dunce? Go on!'

'I'm a poor man,' **the** Hatter went on, 'and most things twinkled after that--only **the** March Hare said--'

'I didn't!' **the** March Hare interrupted in a great hurry.

'You did!' said **the** Hatter.

'I deny it!' said **the** March Hare.

'He denies it,' said **the** King: 'leave out that part.'

'Well, at any rate, **the** Dormouse said--' **the** Hatter went on, looking anxiously round to see if he would deny it too: but **the** Dormouse denied nothing, being fast asleep.

We can see that the overuse of *the* is connected to the overwhelming prevalence of general nouns instead of names in *AW* (see Section 5.1), each of which is preceded by the definite article. This leads to a far greater occurrence of *the* than we would meet with in a text where characters bear proper names.

Off has several functions in the text, as in English in general: it appears as a preposition, an adverb or in the idiomatic phrase *on and off*, etc., mostly in the meaning "away from". Likewise, it is part of the Queen's favourite phrase, "Off with his/her head", which clusters heavily in the final part of the text (see Section 5.7.8). Its most common left-side collocates are verbs of movement, such as *left, walked, hurried* or *ran*, although the sense of *leave off*

doesn't have anything to do with actual movement, but is used in the sense "stop". The following concordance lines show various verbs of movement used with *off* in the text.

d the King eagerly, and he **hurried** off. Alice thought she might as well go
for them, and then quietly **marched** off after the others. 'Are their heads
On various pretexts they all **moved** off, and Alice was soon left alone.
ing late.' So Alice got up and **ran** off, thinking while she ran, as well she
you his history,' As they **walked** off together, Alice heard the King say i
er here.' And the executioner **went** off like an arrow. The Cat's head bega

Down is used in a similar way, as part of phrasal verbs or as a preposition and part of idiomatic phrases (*up and down*). Its most frequent collocate in the text is *sat*, followed by *looked, settled* or *put*. Most of the verbs it follows again denote some kind of movement (*trickle, tumble, fly, flutter*). As with *off*, which was used in the repetitive cluster "Off with his/her head", *down* is used in a three-word cluster "Down, down, down" as much as four times. This is because this phrase is used repetitively to describe Alice's movement through the rabbit-hole.

nder the hedge. In another moment **down went** Alice after it, never once con
some way, and then **dipped** suddenly **down**, so suddenly that Alice had not a m
f before she found herself **falling down** a very deep well. Either the well
rying every door, she **walked** sadly **down** the middle, wondering how she was e
larger than a rat-hole: she **knelt down** and looked along the passage into t
e it written up somewhere.' **Down, down, down**. There was nothing else to do
that all?' said Alice, **swallowing down** her anger as well as she could. 'No

The last keyword of this group is *about*, whose grammatical functions are similar to *down* and *off*, although it is also used as a synonym of the words "around" and "approximately" and in idiomatic expressions such as "Don't be all day about it!" The use of *about* in the sense "around" is the most frequent in the text, used with verbs of movement such as *hunt, walk, swim, splash, sprawl*, etc.

had become of it; so, after **hunting** all about for it, he was obliged to wri
her childhood: and how she would **gather** about her other little children, an
fter a minute or two, they began **moving** about again, and Alice heard the Ra
ce of thunder, and people began **running** about in all directions, tumbling u
sort it was) **scratching** and **scrambling** about in the chimney close above he
Just then she heard something **splashing** about in the pool a little way off,
owd below, and there they lay **sprawling** about, reminding her very much of a

We can see that the prepositions *off, down* and *about* are all used predominantly with verbs of movement, and assume several functions in the text. *Off* and *down* are also part of

repetitive clusters, which increase their frequency in the text. All of them appear in idiomatic expressions.

5.7. Phraseology

5.7.1. 4-Word Clusters in *Alice in Wonderland*

One of the reasons why words in general appear key in a text is their frequent repetition. It is a logical step to follow the analysis of isolated words by an examination of the text's clusters (larger patterns of variable length which repeat in the text). Below, I have divided *AW*'s clusters and other larger structures of a repetitive nature into five groups, depending on various formal characteristics. Some of them have surfaced already in the keyword analysis (most of the members of the a), b) and c) categories). Sections c), d) and e) were collected manually, but especially c) would surface in a corpus analysis of 2-word clusters as well. This was not performed here, as I have focused on larger phrasal structures. Section d) features pairs of individual words which follow one another in the text immediately and occur only once. Section e) shows reversed phrases, also occurring only once in the text. As clusters in computer-assisted research include patters occurring more than once, they can be found only in a) to c) here. The last two sections show pairs of a similar pattern which are not classified as clusters and not studied further in the present analysis. They are given here for the sake of completeness.

All of the structures below point to the centrality of various forms of larger-scale repetition in *AW*. The motivation behind their occurrence is different. Whereas the pairs in d) and e) and clusters in b) are no doubt intentional on the part of Carroll, and are easily spotted without the need of software as they occur within such short span of text, it is hard to say whether recurrent phrases regarding tone and voice in a) are intentional (an exception here might be *Off with his/ her head,* and *And the moral of that is …*). They are dispersed through the whole text and are not so obvious. Clusters in c) are a general property of language, where recurrence is a matter partly of choice and partly of linguistic convention.

a) clusters with variable members: e.g. *in a tone of *, in a * voice, Off with * head! And the moral of that is ** (see Section 5.2).
b) unchangeable clusters (appearing in numerous songs and habitual phrases of *AW),* e.g. *Soup of the evening, beautiful soup; Will you, won't you, will you, won't you, will you*

join the dance?; Do cats eat bats?; Consider your verdict, etc. (see Sections 5.2, 5.3 and 5.4).

c) clusters common to fictional writing and dialogue: *said X, X said, X thought, I wonder, it seems, you know,* etc.

d) reduplicative pairs: *Pat, Pat!; The Duchess! The Duchess!; Dear, dear!; Down, down, down.; I don't like it yer honour, at all, at all!; Curiouser and curiouser!; Which way, which way?*

e) reversed pairs: *I see what I eat - I eat what I see, I like what I get - I get what I like, I breathe when I sleep - I sleep when I breathe, For the Duchess. An invitation from the Queen to play croquet. - From the Queen. An invitation for the Duchess to play croquet.'*

The computer-assisted cluster analysis serves the purpose of bringing to attention those repetitive phrases which are not immediately noticeable in the text. Their length and cut-off point are largely a concern of the researcher's discretion. Generally, the larger a corpus is, the more repetition it is likely to include, and cluster size and cut-off point values can be set quite high. With a short corpus such as our present one, it is not possible to expect numerous occurrences of five word clusters, such as Mahlberg (2012) found in her analysis of clusters in all of Dickens's oeuvre. The cluster size I consider here is four words (terminating at a final stop, but not at a comma) of a minimum frequency of four. The software used for their collection is *Wordsmith 6.0* (2015). For a complete list of collected clusters see Figure 7 in the appendix.

5.7.2. Speech clusters

The most numerous cluster group are, predictably, repetitive phrases of the type "said X (to X)" (see Sections 5.2, 5.3 and 5.4). The list is headed by *said the Mock Turtle* with a frequency of 19 occurrences. It is closely followed by *she said to herself,* which is of course used in reference to Alice. The tendency of the verb *said* to be followed by the prepositional phrase of the type *in a X tone,* which was discussed in Section 5.2, is borne out here by the cluster *said Alice in a*.

In her analysis of Dickens's fiction, Mahlberg (2012: 143-145) discusses clusters of speech, commenting on the fact that they are frequently followed by an indication of manner. The same occurs in *AW*. Clusters of this type play a major role in characterization. *Said the Mock*

Turtle, for instance, is followed by such expressions as *in a deep, hollow tone; at last, with a deep sigh; with a sigh*, but also *in a tone of great relief* or *angrily. The Mock Turtle said* cluster is followed by *with a sigh*. Not only can these expressions help us observe the changes in plot (the Mock Turtle starts off in the story as a melancholy character, who, however, waxes angry or relieved on the subject of school learning) but they can also point to a character's primary characteristic. In the case of the Mock Turtle, it is his association with sadness and sighing which prevails in the text.[42]

The March Hare said cluster is followed by expressions such as *in an encouraging tone, to Alice, very earnestly* (both used when he is prompting Alice to help herself to food) or *to itself in a whisper* (in reference to the anticipated dinner). We can see that he is characterized in a completely different light than the Mock Turtle, and appears in the text as a rather impatient, though good-natured creature. *Said the White Rabbit* is followed by *jumping up in a great hurry*, pointing to his important trait: always being in a hurry. The cluster *said the King and* is preceded in the text by such commands as *You may go, Give your evidence, Call the first witness* or *Don't be impertinent*. These point not only to the centrality of the King's role as a judge in the trial, but also to the fact that he gives rather peremptory commands. *The King said to* cluster, on the other hand, highlights whom the King addresses most in the text. These are mostly people connected with the trial, like the Hatter or the jury.

We can see that the Mock Turtle is a melancholy, though easily provoked character, while March Hare is characterized as good-natured, timid creature. The clusters help to define the Rabbit as an impatient animal, while the King appears in connection to the trial and to his role as a stern judge there.

Let us now look at another highborn character, the Duchess, and the cluster *said the Duchess and*. In this cluster, she is characterized by her abrupt manner of speech and occasional rudeness. However, she is also fond of agreeing with Alice in conversation just to be able to pass on her proverbs. The following excerpts show all of these aspects.

Excerpt 9: Alice in Wonderland (Gardner, 2001: 62- 63, 95-96.)

'It's a Cheshire cat,' said the Duchess, 'and that's why. **Pig!**'

[42] Gardner (2001: 112-113) writes that one possible reason for this dismal characterization of the Mock Turtle could be the fact that "marine turtles often appear to weep copiously- especially females, when they make nocturnal egg-laying visits to the shore."

She said the last word **with such sudden violence** that Alice quite jumped.

'**You don't know much**,' said the Duchess; 'and that's a fact.'

'"**Tis so,**' *said the Duchess*: 'and the moral of that is--"Oh, 'tis love, 'tis love, that makes the world go round!"'

'**I quite agree with you,**' *said the Duchess*; 'and the moral of that is--"Be what you would seem to be" (…)

The cluster *she said to herself*, on the other hand, points to the importance of thoughtful reflection in the text. It is, of course, used in connection to Alice and her musings about the events of her adventure. *Said Alice in a* cluster is also connected to her. It functions in the same way as phrases of the same pattern do for the other characters: to mediate her personality to the reader. Expressions such as *in a melancholy (soothing, coaxing) tone, in a tone of delight* or *in a tone of great surprise* position Alice as an empathic and sensitive heroine, while e.g. *in a great hurry to change the subject* point to her polite upbringing. This is because it is used when Alice has made a conversational blunder or when she sees the conversation going awry.[43] The clusters *as she said this* and *she said this she* form an identical cluster which indicates how dialogue and plot are structured. These clusters are followed by verbs of seeing such as *she looked (up, down), she noticed* and verbs of movement (*she came upon*). They mediate the relationship between thoughts and action in the text. The cluster *so she went on* points to the uncooperative nature of dialogue in Wonderland. It was discussed in sufficient detail in Section 5.4.

5.7.3. Clusters modifying the verbs of saying

There are four recurrent clusters in the text related to the manner of speaking (which I have discussed in Section 5.2 and mentioned also in this section above). These include *in a tone of* (including *a tone of great*), *in an offended tone* and *in a low voice*. If we examine the relations between *in a tone of* and *a tone of great*, we can see that *great* is in fact the most frequent collocate of *tone*. *Great* in these clusters is connected with the words *surprise, curiosity, dismay* and *relief*, expressing a wide range of emotions intensified by it. Other collocates of the cluster *in a tone of* include *delight* and *the deepest contempt*, completing the plethora of emotions experienced by various characters at various moments in the story. If we add to this group the related cluster *in an offended tone* (occurring four times in the text), we find out that there are all of ten occurrences of the word *tone* in the four-word cluster group (and doubtless more besides), making it a very frequent cluster in *AW* indeed. The cluster *in a low voice*

[43] Alice's politeness also surfaces in the cluster *I beg your pardon,* used by her four times in the text.

indicates either secrecy or deference. Its counterpart is the cluster *at the top of* (X's voice) which is discussed further below.

y head's free at last!' said Alice **in a tone of delight**, which changed into
ely story indeed!' said the Pigeon **in a tone of the deepest contempt**. 'I've
hat are they made of?' Alice asked **in a tone of great curiosity**. 'Soles an
Wouldn't it really?' said Alice in **a tone of great surprise**. 'Of course no
BEG your pardon!' she exclaimed in **a tone of great dismay**, and began pickin
I was going to say,' said the Dodo **in an offended tone**, 'was, that the best
ve had nothing yet,' Alice replied **in an offended tone**, 'so I can't take mo
at I say,' the Mock Turtle replied **in an offended tone**. And the Gryphon add
you like the Queen?' said the Cat **in a low voice**. 'Not at all,' said Alic
together, Alice heard the King say **in a low voice**, to the company generally
usly at the White Rabbit, who said **in a low voice**, 'Your Majesty must cross

These clusters serve as an indicator of the manner of speaking, which it is necessary to highlight in the written text, since there is no possibility of actually hearing the speakers. As such, they perform the important role of dramatizing written dialogue.

5.7.4. Clusters of temporal and spatial relations

Temporal and spatial relations, which are expressed in *AW* by four-word clusters *a minute or two* (including the cluster *for a minute or* (two)), *every now and then*, as well as *at the bottom of*, were recognized as an important cluster category in the works of both Fischer-Starcke (2010: 118) and Mahlberg (2012: 67). It is nevertheless important to highlight that these are not just features of written texts, but of language in general. The two collocates of *at the bottom of* are *well* and *sea*, referring to Alice's conversations in *A Mad Tea Party* and with the Mock Turtle respectively. The cluster *at the top of* seems at first like a logical opposite of this spatial relation, but its most frequent collocate is *voice*, indicating rather intensity of utterance than physical place. On the other hand, it is also used in the phrase *at the top of his* (the Frog-footman's) *head*, where it does indicate place. The cluster *the end of the* is used to describe both spatial (*at the end of the bill, with the end of the tail*) and temporal (*before the end of the trial, for the end of the song*) relations. Two other clusters can be grouped in this category: *and when she had*, which points to the temporal relations of Alice's actions, and *she came upon a*, which describes her spatial movements in the text. Clusters indicating time and spatial relations are numerous in the text. Many of them are fixed expressions which help the reader orientate in the plot.

h. CHAPTER VI. Pig and Pepper **For a minute or two** she stood looking at
a well?' The Dormouse again took **a minute or two** to think about it, and t
h at first, but, after watching it **a minute or two**, she made it out to be a
constant howling and sneezing, and **every now and then** a great crash, as if

looking at it uneasily, shaking it **every now and then**, and holding it to hi
Lacie, and Tillie; and they lived **at the bottom of** a well--' 'What did th
self; 'his eyes are so VERY nearly **at the top of** his head. But at any rate
t hurried off, without waiting for **the end of the** song. 'What trial is it?
ear they should forget them before **the end of the** trial.' 'Stupid things!'
ld not help bursting out laughing: **and when she had** got its head down, and
you have just been reading about; **and when she had** finished, her sister ki
s ever to get out again. Suddenly **she came upon a** little three-legged tabl
However, on the second time round, **she came upon a** low curtain she had not

5.7.5. Clusters of the "as…as" type

As well as she and *well as she could* are two four-word clusters which form one five-word cluster, *as well as she could*. These clusters partially overlap with the third member of this category: *as she could for*. All of these relate to Alice and her ability to cope with a difficulty, be it a difficulty with understanding, memory, anger or physical actions such as sneezing, crouching and picking up something. Together, they indicate that Alice gets into a difficult situation of one kind or another several times in the text and always tries to overcome it to the best of her ability (see the comments on "did not like" below).

e was **considering** in her own mind (**as well as she could**, for the hot day ma
that soup!' Alice **said to herself**, **as well as she could for sneezing**. Ther
ot like to be rude, so **she bore it as well as she could**. 'The game's going
id Alice, and she told her sister, **as well as she could remember** them, all
d Alice, **swallowing down her anger as well as she could**. 'No,' said the Ca
lice **crouched down** among the trees **as well as she could**, for the neck kept
y, and began **picking them up again as quickly as she could**, for the acciden

5.7.6. Clusters indicating perception and attitude: "seemed" and "like" type

The clusters *there seemed to be* and *seemed to be no* are fully overlapping. Their collocates include *chance* (of) and *use* (in). They indicate the role of individual perception and expectation in the text, as well as the fact that these perceptions and expectations are frequently negative. Another group of clusters is those which include the word *like (she did not like, like the look of, did not like to, I should like to*). These clusters indicate different aspects of Alice's behaviour in the text, for instance her politeness (*she did not like to* be rude, drop the jar), bashfulness (*she did not like to* go nearer), apprehension (she did not *like the look of* things at all) and resentment (Alice *did not like to* be told so). As we can see, the range of emotions is rather wide. Interestingly, all of these clusters either contain or are preceded by a negative. There are many more aspects of her adventure that Alice does not like, or that appear in a negative light, than those she does like. This would point to the fact that *AW* is primarily a story about negotiation and survival in an alien space.[44]

[44] Throughout the text, Alice continuously negotiates her meaning with the other characters, while she struggles to maintain some order in her conversations with the Wonderland creatures. Her survival in the story equals her

gs indeed were really impossible. There seemed to be **no** use in waiting by
mong the distant green leaves. As there seemed to be **no** chance of getting
t disappointment it was empty: she did **not** like to drop the jar for fear of
ow.' It was, no doubt: only Alice did **not** like to be told so. 'It's really
 It was so large a house, that she did **not** like to go nearer till she had n
having missed their turns, and she did **not** like the look of things at all,
mfortably sharp chin. However, she did **not** like to be rude, so she bore it

5.7.7. Song clusters

A prominent group of clusters, which will not be discussed here as it was discussed already in Section 5.2, are song clusters, consisting of repetitive phrases (which form part of even bigger clusters) such as *will you won't you, won't you will you, you will you won't, you won't you will, would not could not, you join the dance* (from the Lobster Quadrille) and *beau ootiful soo oop* (from the "Turtle Soup" song). The appearance of these clusters is expectable, since all songs commonly feature repetitive phrasal units.

ill you come and join the dance? **Will you, won't you, will you,** won't you
 the dance? Will you, won't you, **will you, won't you,** will you join the d
u, **won't you, will you, won't you, will you** join the dance? Will you, won'
n't you, will you join the dance? **Will you, won't you, will you, won't you**
 not, would not join the dance. **Would not, could not,** would not, could
the dance. **Would not, could not, would not, could not, could not** join the
Soup! Soup of the evening, beautiful Soup! **Beau--ootiful Soo--oop**
l Soup? Pennyworth only of beautiful Soup? **Beau--ootiful Soo--oop**

5.7.8. Personalized phrase clusters

Clusters which are largely predictive from the text are those which are connected to commonly occurring phrases repeated by various characters. We can talk of them being a sort of motto for that given person. One of them is the Duchess's phrase *And the moral of that is*, appearing in the text all of six times (it was mentioned in Section 5.2 and earlier here). The concrete realizations of this phrase in clusters are *and the moral of*, *the moral of that* and *moral of that is*. This cluster emphasizes the role of repetition, but also nonsense in *AW*, since all the proverbs of the Duchess are entirely nonsensical and unrelated to the topic of the conversation. Other repetitive phrases are the clusters *off with his head* and *off with her head* (see Section 5.6). They are a favourite of the Queen's. These clusters support characterization: their frequent repetition and the fact that they appear in a very short span of the text makes them an effective means of character associations.

---

pragmatic ability to gain information from them. She is almost run over by a puppy and the Duchess and the Queen both say that her head should be cut off. In the last chapter, when she is threatened with beheading, she escapes only by finally asserting her authority.

```
lk. I can't tell you just now what the moral of that is, but I shall remem
 ''Tis so,' said the Duchess: 'and the moral of that is--"Oh, 'tis love, 't
mingoes and mustard both bite. And the moral of that is--"Birds of a feathe
 large mustard-mine near here. And the moral of that is--"The more there is
went stamping about, and shouting 'Off with his head!' or 'Off with her hea
all difficulties, great or small. 'Off with his head!' she said, without ev
oment like a wild beast, screamed 'Off with her head! Off--'  'Nonsense!' s
 shouting 'Off with his head!' or 'Off with her head!' about once in a minu
```

5.8. Afterword: the pragmatics of Wonderland dialogue

Throughout this analysis, I have mentioned several times that dialogue forms a large part of the text. Interactions between characters determine the actions of its participants, and there are several chapters characterized mainly by large-scale presence of dialogue (*A Mad Tea Party, The Mock Turtle's Story*, etc.). Wonderland conversations are governed not by the rules (of politeness, logic, language) that Alice knows, but by the nonsense rules of the world she has entered. They are overwhelmingly uncooperative and display features of "*under*politeness" (Leech, 2008: 124), disjointedness and disregard for conventional conversational maxims. The language play of what to take seriously in dialogue and what not to, and the power structures related to politeness and success of interactions, are all explained by Lakoff (1993: 370) as a child's learning experience. Simply said, the rules Alice is just in the progress of learning outside Wonderland are here put into question. Alice's effort to abide by logic and politeness is frequently thwarted by the effort of the other characters to do precisely the opposite. Lakoff (1993: 379) labels this type of discourse as "non-responsive", although he adds that the nonsense it includes blends with perfectly ordinary and meaningful conversation. A frequent feature of it is "power play", not so much over the superiority of the characters themselves in conversation, but over language itself (Lakoff, 1993: 382).

As this is a primarily corpus stylistic perspective, this section is to be taken as an afterword and will be brief and summary. The prevalence of verbs of saying and the manner of saying them, as well as the extensive presence of dialogue in *AW*, leads logically to an inclusion of this pragmatically orientated part at the conclusion of our analysis. It is also based on the fact that pragmatic aspects of Wonderland conversations are an important stylistic feature of the text. I shall take selected excerpts of Wonderland dialogue and examine them in connection to the maxims of the Cooperative Principle and the Maxims of Politeness, resembling Leech's (2008: 118-135) analysis of G.B. Shaw's *You Never Can Tell*. In doing so, I will be able to focus on character interactions which are so important in the text in more detail. The concrete excerpts were chosen in order to include a varied number of Wonderland characters and to

represent several conversation and politeness maxims. Apart from these criteria, the selection was done at random. The maxims used for the analysis are described in detail in Section 4.1.2. Only these shall be used for an explanation of the selected excerpts since they are sufficient to our purpose.

While is it of course impossible to uphold all maxims at once in conversation, since the use of some excludes the use of others, the excerpt below (Excerpt 10) displays several occasions of blatant maxim violation. What is strange about it, and is truly unique of Carroll's writing of *AW*, is that the Wonderland characters generally do not even try to abide by any politeness rules.[45] When they do, the occasion seems rather incidental than otherwise. The excerpt includes the conversation of the March Hare, the Hatter and Alice at the tea party, and the concrete maxims broken in the conversation include the maxim of tact, quantity and relevance. Sympathy and agreement maxims are obviously not observed in the conversation (when the Hatter tells Alice to cut her hair and when she tells him off for it), which is only logical to expect in any situation involving conflict. The only maxim upheld is the generosity maxim, when the March Hare offers Alice refreshment. Even that is warped, however, by the fact that he offers something that he knows is not actually there.

Excerpt 10: Alice in Wonderland (Gardner, 2001: 72-73.)

'Have some wine,' the March Hare said in an encouraging tone.

Alice looked all round the table, but there was nothing on it but tea.

'I don't see any wine,' she remarked.

'There isn't any,' said the March Hare.

'Then it wasn't very civil of you to offer it,' said Alice angrily.

'It wasn't very civil of you to sit down without being invited,' said the March Hare.

'I didn't know it was YOUR table,' said Alice; 'it's laid for a great many more than three.'

'Your hair wants cutting,' said the Hatter. He had been looking at Alice for some time with great curiosity, and this was his first speech.

'You should learn not to make personal remarks,' Alice said with some severity; 'it's very rude.'

If we have a look at how the verbs of saying are modified, we are provided with a sufficient clue to the progress of this conversation. The modifications range from "in an encouraging tone" to "angrily" and "with some severity". As I said above, the interaction starts off politely

[45] An exception might be the Cheshire Cat, who is unusually cooperative when Alice asks it for directions and displays sympathy when he asks her how she is getting on and how she likes the Queen at the game of croquet. He is also the only Wonderland creature who is called her "friend" (see Gardner, 2001: 90-91).

by the March Hare offering Alice wine. When she points out its absence, however, the March Hare does not react with an apology for having offered it when there was not any, but employs the maxim of quantity in a direct statement of fact, "There isn't any". When Alice points out the impoliteness of such behaviour, the March Hare bluntly breaks the tact maxim to point out that she in fact was not invited. The Hatter then follows by another blunt conversational blunder, breaking the maxim of relevance (and of course tact) by saying that Alice's hair is too long. At the end of the conversation, Alice upholds the quality maxim at the expense of tact and says what she believes to be true: the Hatter should not make personal remarks. A similar thing happens in the dialogue below, which includes Alice and the Caterpillar, and in which the maxims of tact, quantity, sympathy and agreement are once again broken. While Alice tries to behave politely throughout the conversation, her every attempt is undercut by the Caterpillar's curt and contrary remarks.

Excerpt 11: Alice in Wonderland (Gardner, 2001: 49-50.)

'Who are YOU?' said the Caterpillar.

This was not an encouraging opening for a conversation. Alice replied, rather shyly, 'I--I hardly know, sir, just at present--at least I know who I WAS when I got up this morning, but I think I must have been changed several times since then.'

'What do you mean by that?' said the Caterpillar sternly. 'Explain yourself!'

'I can't explain MYSELF, I'm afraid, sir' said Alice, 'because I'm not myself, you see.'

'I don't see,' said the Caterpillar.

'I'm afraid I can't put it more clearly,' Alice replied very politely, 'for I can't understand it myself to begin with; and being so many different sizes in a day is very confusing.'

'It isn't,' said the Caterpillar.

'Well, perhaps you haven't found it so yet,' said Alice; 'but when you have to turn into a chrysalis--you will some day, you know--and then after that into a butterfly, I should think you'll feel it a little queer, won't you?'

'Not a bit,' said the Caterpillar.

'Well, perhaps your feelings may be different,' said Alice; 'all I know is, it would feel very queer to ME.'

'You!' said the Caterpillar contemptuously. 'Who are YOU?'

Which brought them back again to the beginning of the conversation.

The Caterpillar starts off rudely by demanding to know who Alice is (notice the stress on "you"). Alice is very polite in her answer (which is characterized by her hesitation and verboseness, but also in her use of the word "sir"). Her reply, however, breaks the maxim of quantity by being a bit too (little) informative. The Caterpillar breaks the tact and agreement maxims again by his next sentence, "What do you mean by that? (…) Explain yourself!" This

is a demand that Alice correct her previous breach of the quantity maxim and be adequately informative. In her next sentence, Alice takes the Caterpillar's "yourself" literally, again at the expense of the quantity maxim (her reply contains very little actual information). The Caterpillar's reply to her next utterance seems very untactful, but he in fact does just what Alice did a moment ago: takes her "you see" literally.[46] One of the reasons why he might be so confused by it is that it is in itself a politeness phrase fostering mutual agreement. As the Caterpillar knows nothing of politeness, he cannot be expected to react appropriately to it. Alice's next sentence again has features of politeness, such as the apologetic "I'm afraid". The Caterpillar, however, once again disagrees with what she says, breaking the agreement maxim. His short, decisive, contrary replies to Alice are designed to be especially uncooperative. Alice, however, perseveres in being polite, as attested by her long replies and expressions such as "perhaps", "you know", "I should think" or "won't you". As in the previous conversation, Alice assumes the role of an outsider to the contrary rules of Wonderland dialogue. While she struggles to uphold the rules she has learned, the creatures do their best at being blunt, direct and contrary. The adverbs following the verbs of saying are again revelatory. While "Alice said" is followed by "rather shyly" or "very politely", the Caterpillar speaks "sternly" and "contemptuously". These expressions serve to demarcate the tenor of the conversation even more clearly.

Alice's movements in Wonderland are heavily influenced by the outcome of her conversations with the creatures that she meets. A common outcome is that either she (e.g. at the tea party, in the *Pig and Pepper* chapter) or some of the other characters (the birds at the pool of tears, the Mouse) escape the conversation out of fear or offence.[47]

Excerpt 12: Alice in Wonderland (Gardner, 2001: 80, 36.)

'Really, now you ask me,' said Alice, very much confused, 'I don't think--'

'Then you shouldn't talk,' said the Hatter.

This piece of rudeness was more than Alice could bear: **she got up in great disgust, and walked off.**

(…)

'A knot!' said Alice, always ready to make herself useful, and looking anxiously about her. 'Oh, do let me help to undo it!'

[46] As Elizabeth Troesch says in *Alice Beyond Wonderland* (Leach & Hollingsworth, 2009: 45), "the Wonderland and Looking-Glass World creatures' tendency toward literal interpretation is what causes most of the breakdowns in communication between themselves and Alice."

[47] Lakoff (1993: 379) sees Carroll's communicative interactions as games. If we accept this view, we can see a conversation participant who retains power at the end of a dialogue as a game winner, while the one who is in the submissive position throughout or exits the conversation as a loser.

'I shall do nothing of the sort,' **said the Mouse, getting up and walking away**. 'You insult me by talking such nonsense!

If there are more characters involved in a conversation, there are generally several interruptions and conversation is rapidly cut off (e.g. in *The Lobster Quadrille*, or when the Gryphon interrupts the Mock Turtle at the end of *The Mock Turtle's Story,* or when both Alice and the March Hare break off the Hatter's account of the origins of the mad tea party in chapter VII, etc.).

Excerpt 13: Alice in Wonderland (Gardner, 2001: 79.)

'No, please go on!' Alice said very humbly; **'I won't interrupt again**. I dare say there may be ONE.'

'One, indeed!' said the Dormouse indignantly. However, he consented to go on. 'And so these three little sisters--they were learning to draw, you know--'

'What did they draw?' **said Alice, quite forgetting her promise**.

'Treacle,' said the Dormouse, without considering at all this time.

'I want a clean cup,' **interrupted** the Hatter: 'let's all move one place on.'

The modesty maxim is continually broken by the Mock Turtle, who is a pedantic know-it-all, calling Alice "a simpleton" (Gardner, 2001: 102) and continuously showing off his "superior" knowledge in the conversation with her. An interesting point to make is that the maxim of quality is never broken in the text. The characters being bluntly truthful at the expense of politeness, there is no reason to lie.

Sometimes, the conversation becomes jarringly different between the characters' rudeness and Alice's politeness[48], such as in the following extract from *The Lobster Quadrille.* While the Mock Turtle utters a peremptory demand (not even addressed directly at the intended recipient, but at an intermediary), Alice expresses her wishes with a "please" and a polite phrase "if the Mock Turtle would be so kind."

Excerpt 14: Alice in Wonderland (Gardner, 2001: 109, 112.)

'I should like to hear her try and repeat something now. Tell her to begin.' He looked at the Gryphon as if he thought it had some kind of authority over Alice.

(…)

'Oh, a song, please, if the Mock Turtle would be so kind,' Alice replied.

[48] Elizabeth Troesch states in *Alice Beyond Wonderland* (Leach & Hollingsworth, 2009: 50) that according to Kathleen Blake (1974), "Alice- with her preoccupation with decorum and rules in game playing- fulfils the role of "adult" in most of her encounters with the inhabitants of Wonderland and the Looking-Glass World."

However, frequently it is Alice herself who breaks the rules of politeness, which happens commonly after she has been goaded by the impoliteness of others. A frequent offense with her is that she interrupts other people's conversations, such as in *The Mock Turtle's Story,* where she interrupts his account of his school days several times, or at the mad tea party, when she jumps into the Dormouse's story about the three sisters living at the bottom of a well. Her other offense is pointing out to the other creatures that she is above them in the food chain (e.g. in her conversation with the Mouse and the birds at the pool of tears or with the Pigeon):

Excerpt 15: Alice in Wonderland (Gardner, 2001: 36, 101.)

'I wish I had our Dinah here, I know I do!' said Alice aloud, addressing nobody in particular. 'She'd soon fetch it back!'

'And who is Dinah, if I might venture to ask the question?' said the Lory.

Alice replied eagerly, for she was always ready to talk about her pet: 'Dinah's our cat. And she's such a capital one for catching mice you can't think! And oh, I wish you could see her after the birds! Why, she'll eat a little bird as soon as look at it!'

The former rudeness might be accounted for by the fact that Alice is too eager to show off her logical thinking or knowledge to others, and also that she has a very enquiring mind. The latter positions her momentarily into the role of an unknowing buffoon for the purpose of comedy. After her conversational blunders at the pool of tears, Alice begins to realize that conversations in Wonderland abide by different rules than those outside of it. Even so, she continues making similar mistakes later on.

Not all Wonderland dialogue is governed by disregard for cooperation. A strangely convoluted instance of a Wonderland character trying to observe the generosity (and tact) maxim happens at the beginning of the text in chapter III. When the animals have finished running the Caucus-race to get dry after swimming in the pool of Alice's tears, the task of giving out prizes is unceremoniously delegated to Alice by the Dodo. When Alice has given away everything in her pockets, the Mouse tactfully observes that she has not got any prize. This tact might have been welcome if it were not for the fact that Alice is simultaneously the giver and the recipient.

The sympathy maxim is unexpectedly upheld when the Duchess meets Alice at the Queen's party and greets her enthusiastically after being extremely hostile to her in the *Pig and Pepper* chapter.

Excerpt 16: Alice in Wonderland (Gardner, 2001: 33, 94.)

'But she must have a prize herself, you know,' said the Mouse.

'Of course,' the Dodo replied very gravely. 'What else have you got in your pocket?' he went on, turning to Alice.

'Only a thimble,' said Alice sadly.

'Hand it over here,' said the Dodo.

Then they all crowded round her once more, while the Dodo solemnly presented the thimble, saying 'We beg your acceptance of this elegant thimble'; and, when it had finished this short speech, they all cheered.

(…)

'You can't think how glad I am to see you again, you dear old thing!' said the Duchess, as she tucked her arm affectionately into Alice's, and they walked off together.

In the following excerpt (Excerpt 17), Alice talks to the gardeners in *The Queen's Croquet-Ground* chapter. As we can see, Alice begins politely with "Would you tell me", and whereas Five and Seven do not cooperate with her, Two explains profusely (with expressions such as "you see, Miss" or "you know") the reasons for repainting the rose bush. He upholds both the maxims of quantity and relation. The verbs of speaking are deferential and shy on both sides: Alice speaks "a little timidly" and the gardener answers "in a low voice". The reason for this politeness might be that the gardeners are in the position of subservience, not freestanding characters. This is actually one of the few truly cooperative conversations Alice has in the course of her adventure.

Excerpt 17: Alice in Wonderland (Gardner, 2001: 84.)

'Would you tell me,' said Alice, a little timidly, 'why you are painting those roses?'

Five and Seven said nothing, but looked at Two.

Two began in a low voice, 'Why the fact is, you see, Miss, this here ought to have been a RED rose-tree, and we put a white one in by mistake; and if the Queen was to find it out, we should all have our heads cut off, you know. So you see, Miss, we're doing our best, afore she comes, to--'

At this moment Five, who had been anxiously looking across the garden, called out 'The Queen! The Queen!' and the three gardeners instantly threw themselves flat upon their faces.

In the last excerpt below (Excerpt 18), we can see that even the Duchess, who has no reason to be civil to Alice, behaves politely in *The Mock Turtle's Story* chapter.[49] She upholds the agreement maxim, agreeing with everything Alice says ("Of course it is", "I quite agree with you") in an effort to pass on her proverbs. On the other hand, she completely violates the maxim of relation (and manner), saying something completely irrelevant to the topic of conversation after the phrase "and the moral of that is". Furthermore, she breaks the maxim of quantity. While the utterance "if you'd like it put more simply" should be followed by a briefer account of what has been said, the reverse is true. The Duchess might agree with Alice and uphold politeness, but she is incapable of following other basic maxims of conversation. On the other hand, Alice also fails to abide by the rules, breaking the tact maxim by not listening to a word the Duchess says.

Excerpt 18: Alice in Wonderland (Gardner, 2001: 96-97.)

'It's a mineral, I THINK,' said Alice.

'Of course it is,' said the Duchess, who seemed ready to agree to everything that Alice said; 'there's a large mustard-mine near here. And the moral of that is--"The more there is of mine, the less there is of yours."'

'Oh, I know!' exclaimed Alice, who had not attended to this last remark, 'it's a vegetable. It doesn't look like one, but it is.'

'I quite agree with you,' said the Duchess; 'and the moral of that is--"Be what you would seem to be"--or if you'd like it put more simply--"Never imagine yourself not to be otherwise than what it might appear to others that what you were or might have been was not otherwise than what you had been would have appeared to them to be otherwise."'

These short excerpts will hopefully serve to give a summary idea of what happens in conversations with the Wonderland characters throughout the text. According to Flescher (1969: 137-138), "conversation, or more precisely, argument, is the essential vehicle of nonsense" in *AW* and it functions on the principle of constant deflection and misinterpretation. The creatures constantly break the cooperative principle and politeness maxims because their pragmatic upbringing is different from Alice's. It is not so much that they do not know the rules; it is more that they do not see why they should abide by them if blunt truth or self-interest is better served by directness. It is however doubtful if they intentionally mean to be

[49] This behaviour is completely different from the one in the *Pig and Pepper* chapter, where the Duchess responds curtly and rudely to Alice's questions about the Cheshire cat. The Duchess's ill-temper in this chapter is explained by Gardner (2001: 62) as the effect of the fiery pepper in the soup. Her proverbial "Oh, 'tis love, 'tis love, that makes the world go round!'" from chapter IX is radically altered in the *Pig and Pepper* chapter to: 'If everybody minded their own business,' (…) 'the world would go round a deal faster than it does' (Gardner, 2001: 63).

rude in their conversations; it is more likely that there is no malicious intention behind their nonsense and that it just happens for no reason, as does Alice's jumbled poetry. There are exceptions to this conduct, of course, and there are instances where Alice herself is at fault, commonly of interrupting somebody's speech, but she normally strives to be polite or apologizes for misconduct. While the creatures frequently disregard cooperativeness and politeness altogether, Alice's faulty turn-taking appears less serious in comparison with their open rudeness.

The general progression of interactions in Wonderland is from agreement to disagreement, with frequent instances of obscurity, lack of relation, too little (or too much) information, and chronic lack of tact. Some adherence to the rules of dialogue is necessary for the conversations to progress at all, but even that progression is often given only by Alice's unusual willingness to participate in it. As said in Inaki and Okita's (2005) article on Alice's role in Wonderland, her behaviour is predominantly passive, where her polite upbringing overrules her assertion of power in the text. The creatures abide by their nonsense rules and force Alice to accept it. In doing so, the conversational power rests on their side. It is in the final chapter that Alice finally asserts her authority and exits her adventure.

Excerpt 19: Alice in Wonderland (Gardner, 2001: 129.)

'Off with her head!' the Queen shouted at the top of her voice. Nobody moved.

'Who cares for you?' said Alice, (she had grown to her full size by this time.) 'You're nothing but a pack of cards!'

6. Conclusion

The goal of this thesis was to identify stylistically prominent patterns in Lewis Carroll's *Alice in Wonderland* (1865) using corpus stylistics approach. I started with an explanation of key stylistic terms, such as the notion of norm, foregrounding, deviance and relevance. Later on, I highlighted the benefits of the corpus approach and focused in detail on two concepts traditionally studied by corpus linguists: keywords and clusters. In the methodological part I discussed key corpus linguistic terms, the role of reference corpora in keyword analysis and cluster length and cut-off point. The fourth chapter focused on *Alice in Wonderland* and its role in the Golden Age of children's writing. In the following sections, the role of nonsense in the text was explored and pragmatic principles connected to Wonderland communicative exchanges were stated. Lastly, the reference corpora and hypothesis were discussed.

The choice of two reference corpora, the corpus of children's fiction and the corpus of contemporary literature, led to comparable results. The number of keywords on the contemporary literature list was slightly higher, which was caused by the bigger size of the corpus, but other than that the differences were minor. This supports Scott's (2006) conclusion that the robustness of keyword analysis is upheld regardless the choice of a reference corpus. On the other hand, the fact that I used reference corpora of a similar genre and similar period probably aided the analysis as a whole.

The two main predecessors of the type of analysis performed here include Bettina Fischer-Starcke (2009, 2010) and Michaela Mahlberg (2007, 2012). Both performed a corpus study of literary fiction, focusing in the first case on keywords and phraseology of *Pride and Prejudice* and *Northanger Abbey*, and in the second case on prominent clusters in Dickens's fiction and their role in characterization. Some of the themes which they identified were found here, too (e.g. verbs of saying, temporal and spatial clusters), leading to the conclusion that these are a staple property of literary language, particularly Victorian literature. Other themes found in the course of the present analysis were unique and therefore indicators of the specificities of *Alice in Wonderland* as opposed to other works of fiction (e.g. the role of *the*, individual repetitive clusters of various characters…).

The practical part of the analysis started with an exploration of the text's keywords, which were sorted into lexical and grammatical domains according to their part of speech. The keyword analysis led logically to a section devoted to four-word clusters of *Alice in Wonderland*, which were again sorted into several categories according to their semantics and content. In these two parts were identified several major characteristics of the text, the most prominent of them being the prevalence of verbs of saying postmodified by a qualifying adverb or prepositional phrase of the type "in – a – adjective – voice/tone". Together with contracted forms of grammatical verbs, this theme led to conclusion that *Alice in Wonderland* contains an unusually large amount of dialogue, a fact which was later explored from a pragmatic point of view in the last section of this analysis. In addition, one of the key factors proved to be different types of character names, which position their bearers on different levels in the power structures of the text. In the cluster section, repetitive phrases found in songs or in the speech of characters were discussed in detail, as well as clusters supported by the keyword analysis, and others.

Food and eating (keywords *eat, bread-and-butter, pepper, soup* or *tarts*), justice (words like *executioner, jury, court, trial, witness,* etc.), personal identity (keywords *large, little* or *size*), as well as clusters related to characterization (see Sections 5.7.2 and 5.7.3 on manner of speaking) and negativity (see Section 5.7.6 on the verb like and its connection to negation) have all surfaced as the major thematic concerns of the text.

Other key concerns of the text included the differentiation between various types of names, which together with a section on personal possessions and manner of speaking broached the theme of characterization and character depiction in the story. The heroine herself was found to be a reflecting, empathic and sensitive one, who assumed a rather timid role in the story. This conclusion was also borne out in the pragmatic section, where selected dialogical passages were explored on the basis of Grice's Maxim and the Maxims of the Cooperative Principle. This section was included on the basis of the high content of dialogue and its unique pragmatic nature in the text. It was discovered that while Wonderland characters are freely uncooperative and disregardful of the rules of felicitous communication, Alice tries to maintain politeness at the cost of her own conversation power.

The analysis combined the benefits of both the quantitative and qualitative approach and illumined some of the most prominent stylistic markers of a much-loved work of children's

literary fiction. The discoveries that I have been able to make on its basis lead me to the conclusion that the corpus exploration of texts plays an important role in the stylistic study of literature, and that it is capable of identifying its key concerns and characteristics.

7. Shrnutí práce v češtině

Náplní této práce je korpusově založená stylistická analýza *Alenky v říši divů* (1865) od Lewise Carrolla. Na úvod práce jsem se zabývala podrobným vysvětlením pojmu styl a rolí normy v jeho určování. Jak je uvedeno v sekci 2.1, norma je relativní koncept představující jazykovou aproximaci a měla by být blízká zkoumanému dílu žánrem i historickým obdobím.

Termíny foregrounding (spojené s českým pojmem *aktualizace*), *deviance* (deviace, ochylka) a *prominence* (prominence) jsou klíčovými v oblasti stylistiky (Leech, 2008: 18). Aktualizace představuje stylisticky vybočující charakteristiky textu, které se v něm mohou stát prominentními. Je nutno dodat, že pouze deviace je statisticky měřitelná, protože aktualizace se týká psychologického efektu, na jehož základě čtenář přisuzuje stylistickému znaku významnost. Zároveň ne každá statistická odchylka aktualizaci tvoří. Princip aktualizace se týká především dvou vlastností textu: paralelismu (obdobných struktur) a deviace (struktur odlišných). Leech (2008) rozlišuje tři typy vybočení (deviace), a to deviaci primární, sekundární a terciární. Tato práce se zabývá deviací sekundární, která se týká odlišnosti textu od konkrétní literární kompozice (v tomto případě od textů dětské literatury a soudobé viktoriánské tvorby).

Sekce 2.2 pojednává o jedné z hlavních výhod korpusového přístupu: statisticky doložitelné objektivitě. Ta se v práci samé nutně snoubí se subjektivním ohodnocením badatele, ale přesto se předpokládá, že subjektivita je nižší. Za jeden z nejvýznačnějších znaků textu se považuje opakování, např. klíčových slov či větších frazeologických jednotek (tzv. *clusters* v korpusové lingvistice, tedy n-gramů). V oblasti klíčových slov se většinou rozlišuje mezi lexikálními jednotkami, které indikují děj, a gramatickými, které indikují strukturu textu. Za limitace korpusového přístupu se považuje např. možná ztráta individuálního přístupu badatele, rozdílné výsledky při použití rozdílných statistických měr, či snadná zaměnitelnost některých slovních druhů (např. podstatných jmen a sloves, užívá-li se značkování, což není případ této práce), která mohou zkreslit počet výskytu jednotlivých slov.

Sekce 2.3 uvádí šest významných autorů, jejichž práce přispěly k rozvoji korpusové stylistiky, tedy zkoumání literárních děl korpusovými technikami. Z nich nejvýznamnějšími pro tuto práci jsou zejména Bettina Fischer-Starcke (2009, 2010) a Michaela Mahlberg (2007, 2012),

které se zabývaly *klíčovými slovy* (statisticky signifikatně častěji se vyskytujícími slovy) a frazeologií v dílech Jane Austen a Charlese Dickense.

V sekci 2.4 se zabývám klíčovými slovy, jejich selekcí, hodnotou pravděpodobnosti a statistickými testy a argumentuji pro zařazení jak lexikálních, tak gramatických slov do analýzy, jelikož oboje mohou do studie vnést důležité poznatky. Dále zmiňuji i důležitost tzv. negativních klíčových slov, kam se zařazují slova, která jsou ve zkoumaném korpusu signifikantně málo zastoupená a mohou tedy upozornit na důležité absence v textu (oproti normě, tedy zvolenému referenčnímu korpusu). Nakonec zmiňuji výhody sémantické klasifikace slov, která umožňuje rozpoznání klíčových významových slovních skupin.

Sekce 2.5 se zaměřuje na analýzu n-gramů, tedy *n* po sobě se opakujících slov. Fischer-Starcke (2010) poukazuje na jejich schopnost poukázat na prominentní stylistické znaky textu, které se vyskytují v delších souslovích, na možnost studování koheze a spojení slov s kolokacemi a koligacemi. N-gramy nacházejí své využití jak v korpusově lingvistických studiích obecně, tak při zkoumání literárních děl, jako tomu je např. u již zmiňované Mahlberg (2007, 2010). Ta uvádí jako jejich hlavní vlastnost, že poukazují na lokální textové funkce. V literatuře nutně dochází k překryvu jednotlivých sémantických skupin n-gramů, protože kategorie jako časové a prostorové značení a kvalitativní/kvantitativní vztahy se ve víceslovných spojeních vyskytují běžně téměř v jakékoliv beletrii.

V metodologické části, začínající úsekem 3.1, se nejprve věnuji vlastnostem softwaru Wordsmith (Scott, 2012), který je v práci využíván. Termíny jako konkordance (blízký kontext vyhledávaného slova (*node*)), *wordlist* (seznam slov vyskytujících se v daném korpusu) či kolokace (obvyklý spolu výskyt slov) a její rozsah jsou vysvětleny. Dále se krátce zabývám termínem konotace, označujícím přičlenění významu slova na základě kulturních, individuálních a historických zkušeností čtenáře. Také je vyzdvižena důležitost zkoumání rozložení klíčových slov v textu (*keyword plot*). Sekce končí odůvodněním, proč se tato práce zabývá ve frazeologické sekci čtyř slovnými n-gramy, k čemuž dochází z důvodu reprezentovatelnosti a zároveň ne příliš velkého počtu těchto frází v *Alence*. Nakonec zmiňuji svůj záměr se zabývat konverzační pragmatikou vybraných částí textu na základě Griceových (1975) principů kooperace a Leechových (2008) zdvořilostních principů.

Sekce 3.2.1 poukazuje na důležitost výběru referenčního korpusu v lingvistické analýze textu. Závěrem argumentace je, že je nejrozumnějším řešením vybírat referenční korpus na základě jeho žánrové, historické a typové podobnosti vůči studovanému korpusu, přičemž referenční korpus bývá větší. Dalším sporným tématem je množství klíčových slov, která by měla být obsažena v analýze. Zatímco Gabrielatos a Marchi (2011) uvádějí, že je obvyklé využít prvních sta slov, tato práce využívá všech položek v seznamu, jelikož ten výrazně nepřesáhl tuto v kvalitativní analýze zvládnutelnou hranici.

Sekce 3.2.2 se zabývá délkou a počtem opakování n-gramů v textu, kde z důvodu menšího rozsahu a repetitivnosti *Alenky v říši divů* volím zahrnutí čtyř slovných jednotek, jejichž celkový počet v textu dosahuje 55 (n-gram se v textu opakuje alespoň čtyřikrát).

Čtvrtá část práce se již konkrétně zabývá materiálem využitým k analýze: *Alenkou v říši divů (Alice in Wonderland)*. Nejprve je nastíněn vývoj dětské literatury před jejím sepsáním a dále jsou charakterizováni hlavní představitelé tzv. zlatého věku dětské literatury (Golden Age of children's literature (Knowles & Malmkjaer, 1995: 16; Hunt, 1994)), přičemž je popsán jejich vztah k *Alence*, stejně jako jejich odlišnost od ní. V sekci 4.1.2 je popsána role „nesmyslu" v *Alence*, který zde funguje na principu pevně stanovených pravidel. Jeho účelem v textu je jazyková hra a zkoumání arbitrárních pravidel konverzačních interakcí včetně vztahů mezi výrazem a tím, co značí (které jsou v mnoha případech značně pozměněny). Jako taková se *Alenka* řadí k prvním dílům moderní dětské literatury.

Na konci sekce popisuji pravidla konverzačních interakcí podle Griceových principů kooperativnosti a Leechových principů zdvořilosti, přičemž naznačuji, že dialog je v *Alence* stavěn na principu narušení pragmatických pravidel. Zde je důležité uvést, že všechny rozhovory v *Alence* nejsou naprosto nekooperativní, ovšem valná většina se tímto směrem ubírá. V závěru zmiňuji článek od Inakiho a Okity (2005), kteří se zabývali rolemi Alenky a jejich vývojem v *Říši divů* a *Za zrcadlem* a došli k závěru, že v prvním díle příběhu se Alenka jeví jako pasivní a neprůbojná hrdinka.

V oddílu 4.2 popisuji formální náležitosti *Alenky* a svých dvou referenčních korpusů a předkládám předběžné hypotézy.

Sekce 5 je již věnována samotnému výzkumu. V prvních šesti oddílech jsou zkoumána klíčová slova podle slovních druhů a jejich sémantiky, sekce sedmá se zabývá n-gramy a poslední sekce je věnována pragmatice.

V oddílu 5.1 se zabývám jmény postav. Nejprve vysvětluji rozdíl mezi tzv. hrubou frekvencí jejich výskytu a hodnotou „klíčovosti" (*keyness value*) a rozebírám některé rozdíly mezi výsledky porovnání *Alenky* s korpusem dětské literatury a soudobé viktoriánské literatury. Poté jsou postavy rozděleny do sedmi skupin, a to: A) obecná jména sloužící jako vlastní jména, B) obecná jména sloužící jako vlastní jména s adjektivní premodifikací, C) obyčejná obecná jména, D) jména označující karetní postavy a dvorské funkce, E) jména lidských postav, F) vlastní jména označující postavy, které pocházejí ze světa mimo říši divů a G) jména odkazující ke zvířatům. Zvířecí jména jsou v říši divů nejčastější, což se váže také k charakterizaci knihy jako dětská literatura.

V první kategorii jmen najdeme postavy jako *Gryphon*, *Mouse* či *Lory*. Tato jména slouží zároveň jako indikátor zvířecího druhu a jako vlastní jméno. První kategorie toho má mnoho společného s druhou: liší se pouze úvodním adjektivním prvkem, jako u jmen *White Rabbit* či *Mock Turtle*. V třetí kategorii najdeme postavy jako *hedhehog, flamingo* či *puppy*, které slouží jako obecná jména méně důležitých postav a nejsou psána s počátečním velkým písmenem. Čtvrtá kategorie je tvořena postavami karet, jako jsou *King, Queen* či *gardeners* a *players*, zatímco pátá obsahuje lidské postavy, jako je *Hatter* nebo *Duchess*. V předposlední kategorii najdeme pouze dvě postavy: Alenku a její kočku *Dinah*, zatímco poslední je tvořena odkazy ke zvířatům, které se v textu neobjevují jako postavy: např. *lobsters* či *porpoise*. Jméno *Alenka* se v těchto kategoriích jeví jako výjimečné, neboť kromě toho, že slouží jako Alenčino vlastní jméno, zároveň neindikuje její druh jako lidská dívenka. Je důležité poznamenat, že zvířecí kategorie v této sekci jsou indikativní, co se týče statusu a role nositelů těchto jmen v *Alence*. Jména psaná s velkým počátečním písmenem se obecně vztahují k důležitým postavám příběhu, zatímco ta s malým písmenem slouží jako indikátor méně důležitých postav a těch zvířat, která se v příběhu vůbec neobjeví.

Sekce 5.2 se zabývá lexikální kategorií podstatných jmen, která jsou rozdělena do skupin jídlo (např. *pepper, tarts*), osobní majetek (*fan, hookah*…), soudnictví (*court, jury, trial*…), písně a verše (*winkle, chorus*…), mluva (*tone, voice*…) a další. Z bližší analýzy těchto slov vyplývá, že jídlo a osobní majetek slouží k charakterizaci postav, ke kterým se unikátně vážou. Téma

soudnictví je důležitým indikátorem děje a odkazuje na prominenci dvou posledních kapitol v textu. Slova vyskytující se v písních a verších se stávají klíčovými díky své neobvyklosti a repetitivnosti, ovšem jejich výskyt v textu je pouze místní. Naopak výrazy týkající se mluvy jsou všudypřítomným znakem Carrollova stylu a váží se ke slovesům mluvení jako jejich doplněk naznačující styl promluvy. V této sekci také uvádím několik tabulek, naznačujících kde v ději se daná slova vyskytují. Tyto tabulky napomáhají v identifikaci daných výrazů jako izolované a naopak pervazivní prvky.

Část 5.3 se zabývá klíčovými přídavnými slovy a příslovci v *Alence*, která obsahují výrazy jako *timidly, anxiously, hastily* a *very, large, little* či *offended*, atd. Uvedená příslovce se váží především k již zmíněnému způsobu promluvy, zatímco přídavná jména evokují motiv změny velikosti (a tím i identity), jimiž Alenka v příběhu prochází, či se váží k několika důležitým motivům v textu: např. motiv nekooperativnosti v konverzaci (*offended*), nesmyslu (*mad, curious*) či soudnictví (*executed*).

Sekce 5.4 je důležitá z toho hlediska, že hovoří o jednom z nejvýznamnějších témat *Alenky v říši divů*: dialogu. Předkládá totiž závěr, že většina sloves obsažených v textu jsou slovesa mluvení, z nichž nejčastější je sloveso *said*. V souvislosti s výrazem *to herself* poukazuje toto sloveso k tématu introspektivnosti hrdinky (vyjádřené rovněž slovesem *think*). Slovesa *began* a *went on* naznačují počátky promluvy a její pokračování. Jejich kolokace poukazují k již zmiňovanému tématu nedodržování konverzačních pravidel, protože vykazují několik případů, kde se zúčastnění buďto neposlouchají, ignorují, či přerušují. Sloveso *know* se ve velké míře vyskytuje ve frázi *you know* a zdůrazňuje tím roli mluvené interakce v textu. Ta je podložena i gramatickými slovesy jako *'m* nebo *doesn't*, která se vyskytují v *Alence* ve zkrácené formě běžně používané pouze v dialogu. Posledním zmíněným slovesem je *eat,* které se váže k tématům jídla a změny podoby, která byla již zmíněna.

Sekce 5.5 uvádí klíčová zájmena v textu, což jsou *she, herself, it* a *its*. Ta dokazují fakt, že Alenka jako hlavní hrdinka má v textu velice význačnou roli, a že dalšími postavami v příběhu jsou zvířata, k nimž tradičně v angličtině přináleží neutrální zájmeno *it*. V této části je ovšem také zmíněno, že zvířecí postavy se často pojí jak s neutrálním, tak s mužským zájmenem, což ukazuje na nesrovnalosti reference v textu. Jedním vysvětlením může být, že jsou vnímány jednak jako zvířata, ale také jako postavy s lidskými vlastnostmi.

Poslední část věnovaná klíčovým slovům se zabývá zbylými gramatickými slovními druhy v *Alence*, jako *the* nebo *off*. Zajímavá je zejména klíčovost členu *the*, který je běžně v angličtině velmi častý. V analýze docházím k závěru, že je to dáno zejména tendencí používat tento člen před jmény naprosté většiny postav (např. *the Queen, the Dodo, the Hatter*, atd.).

Sekce 5.7 se již zabývá n-gramy, které jsou nejprve rozděleny do několika kategorií dle jejich formálních vlastností, a to: A) n-gramy s jedním variabilním členem (např. *in a tone of *, in a * voice*), B) s neměnnými členy (*Do cats eat bats?; Consider your verdict!…*), C) n-gramy, které se běžně vyskytují v literatuře a dialogu (*said X, it seems, you know…*), a D) reduplikativní páry (*The Duchess! The Duchess!; Dear, dear!)* a obrácené páry (*I see what I eat - I eat what I see*). Tyto kategorie byly částečně sestaveny bez pomoci korpusu na základě vlastní analýzy textu. V samotné praktické části se pak již věnuji jen těm, které se objevily na seznamu v softwaru Wordsmith (Scott, 2012).

Část 5.7.2 se zabývá n-gramy týkajícími se mluvení, jako *said the Mock Turtle* či *she said to herself*. Jak je vidno, navazuje na předchozí sekce pojednávající o mluvení a způsobu promluvy. Díky zahrnutí jednotlivých postav v rámci širší, čtyřčlenné frazeologické jednotky, bylo ale možné zkoumat, jak dané n-gramy souvisí s charakterizací postav. The Mock Turtle například nabyl charakterizace jako melancholický pedant, zatímco the March Hare se ukázal jako převážně dobrosrdečná postava a the Rabbit jako uspěchaný „věčně pozdě“. Alenka, na druhou stranu, je vyobrazena jako empatická a citlivá hrdinka.

Část 5.7.3 se znovu zabývá frázemi týkajícími se způsobu mluvení, kde je zjištěno, že nejčastějším členem této kategorie jsou n-gramy obsahující slovo *tone*. Dalšími zajímavými členy jsou n-gramy naznačující hlasitost promluvy, jako *in a low voice* či *at the top of (his voice*). Všechna slovní spojení této skupiny slouží jako indikátory způsobu promluvy v psaném textu, kde čtenář nemá jinou možnost zjistit, jak bylo to či ono řečeno.

Sekce 5.7.4 se věnuje n-gramům vyjadřujícím časové a prostorové vztahy v textu. Jejími členy jsou např. *in a minute or two, every now and then, at the bottom of* nebo *at the end of the*. Tyto frazeologické jednotky jsou přirozenou součástí jazyka a jsou téměř neměnné. V *Alence* slouží k vyjádření časoprostoru hrdinčina dobrodružství a napomáhá čtenáři v orientaci v textu.

Sekce 5.7.5 a 5.7.6 se zabývají n-gramy typu „*as….as“* a „*seemed“* či „*like“*. Zde prvně uvedené se váží k tématu obtíží v textu, do kterých se Alenka několikrát dostane (např. *as well as she could*) a mají přímý vztah k motivům nesnadnosti počínání a následné vytrvalosti, jíž se Alenka jako postava vyznačuje. N-gramy jako *there seemed to be no (chance of, use in)* naznačují prvek negativní perspektivy v textu, zatímco n-gramy typu *she did not like* nebo *did not like to* podporují téma negativního působení vnějších sil na hrdinku, která jim vzdoruje. To se samozřejmě váže také k motivu nekooperativnosti v dialogu.

Poslední dvě sekce (5.7.7 a 5.7.8) odkazují k frázím běžně se vyskytujícím v písních a básních říše divů a opakovaným frázím jednotlivých postav, které přispívají ve velké míře k jejich charakterizaci, jako *and the moral of that is* či *off with his/her head*.

V poslední části práce (5.8) se zabývám dialogovými interakcemi mezi jednotlivými postavami z pragmatického hlediska. Docházím k závěru, že pravidla kooperativnosti i zdvořilosti jsou systematicky porušována, zejména ze strany postav z říše divů, které nedisponují běžnými pragmatickými pravidly. Nejčastěji porušovaná pravidla jsou pravidla kvantity (je řečeno příliš málo či naopak příliš mnoho), relace (daná promluva se nevztahuje k tomu, co bylo řečeno předtím), způsobu (promluva je nejasná či dvojznačná) atd. Také nejsou dodržována pravidla sympatie a souhlasu, ovšem především taktu, který je systematicky narušován. Na druhou stranu dodržovaným pravidlem se jeví pravidlo kvality: postavy jsou příliš přímé na to, aby lhaly. Alenka je oproti tomu v komunikaci ovlivněna pragmatickými pravidly a slušným vychováním reálného světa, a ve většině interakcí se snaží tato pravidla dodržovat na úkor svého postavení v konverzaci, kdy je tím pádem delegován větší prostor ostatním postavám. Na druhou stranu se i ona dopouští konverzačních přestupků, a to zejména co se týče tzv. „turn-taking“, což se projevuje tím, že často skáče jiné postavě do řeči.

Tento popis naznačuje pravdivost tvrzení Inakiho a Okity (2005), kteří uvádí, že role Alenky v průběhu jejího dobrodružství v říši divů je především pasivní. Je ovšem nutno dodat, že v textu jde především o hru s jazykovými a pragmatickými pravidly, která jsou obratně převrácena, spíše než o roli konverzační (nebo i reálné) moci jednotlivých postav v jednom z nejzajímavějších děl dětské literatury. Závěrem tedy řekněme, že pragmatická analýza

interakcí v *Alence v říši divů* přispívá k charakterizaci tohoto textu jako jazykově výjimečného unikátu.

List of references and sources

Altenberg, B. 1998. On the phraseology of spoken English: the evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: theory, analysis and applications*. Oxford: Oxford University Press, 101–122.

Anthony, L. 2014. AntConc (3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/ [Last accessed 11 August 2015].

Archer, D. 2009. *What's in A Word-List?* Farnham: Ashgate.

Baker, P. 2004. Querying keywords : questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics*, 32 (4), 346-359.

Biber, D., Conrad S. & Reppen, R. 2004. *Corpus Linguistics*. Cambridge [u.a.]: Cambridge Univ. Press.

Biber, D. 2006. *University Language: A Corpus-based study of spoken and written registers*. Amsterdam/New York: John Benjamins.

Blake, K. 1974. *Play, Games, and Sport*. Ithaca: Cornell University Press.

Bondi, M. & Scott. M. 2010. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Pub.

Brown, P. & Levinson, S.C. 1987. *Politeness*. Cambridge [Cambridgeshire]: Cambridge University Press.

Carpenter, H. 2009. *Secret Gardens: A Study of the Golden Age of Children's Literature*. Faber & Faber.

Cheng, W., Greaves, C. & Warren, M. 2006. From N-Gram to Skipgram to Concgram. *IJCL* 11 (4), 411-433.

Carroll, L. 1865. Alice in Wonderland. *Project Gutenberg*. Available at https://www.gutenberg.org/cache/epub/19033/pg19033.txt [Last accessed 11 August 2015].

Cohn, D. 1978. *Transparent Minds: Narrative Modes for Presenting Consciousness in Fiction*. Princeton: Princeton UP.

Culpeper, J., Kytö, M. 2002. Lexical Bundles in Early Modern English Dialogues: A Window into the Speech-related Language of the Past. In Teresa Fanego,Belén Méndez-Naya,Elena Seoane (red.), *Sounds, words, texts and change*. Amsterdam/Philadelphia: John Benjamins, 45-63.

Culpeper, J. 2009. Keyness: Words, Parts-Of-Speech and Semantic Categories in the Character-Talk of Shakespeare's Romeo and Juliet. *IJCL* 14 (1), 29-59.

Deignan, A. 2005. *Metaphor and Corpus Linguistics*. Amsterdam: J. Benjamins Pub.

Dušková, L. et al. 2009. *Mluvnice současné angličtiny na pozadí češtiny*. Electronic edition. Ústav anglického jazyka a didaktiky, FF UK. Available at http://www.mluvniceanglictiny.cz [Last accessed 11 August 2015].

Fischer-Starcke, B. 2010. *Corpus Linguistics and the Study of Literature*. London: Continuum.

Fischer-Starcke, B. 2009. Keywords and Frequent Phrases of Jane Austen's Pride and Prejudice: A Corpus-Stylistic Analysis. *IJCL* 14 (4), 492-523.

Flescher, J. 1969. The Language of Nonsense in Alice. *Yale French Studies*, 43, 128-144.

Gabrielatos, C. & Marchi, A. 2011. Keyness: Matching metrics to definitions. Corpus Linguistics in the South: Theoretical-methodological challenges in corpus approaches to discourse studies – and some ways of addressing them. *University of Portsmouth*.

Gardner, M. 2001. *The Annotated Alice*. London: Penguin Books.

Grice, H. P. 1975. *Logic and conversation*. Cole, P., and J.L. Morgan, eds. Speech Acts. New York: Academic Press, 41–58.

Hanks, P. 2013. *Lexical Analysis*. Cambridge, Mass.: The MIT Press.

Hitt, J. 2012. Words on Trial - The New Yorker. *The New Yorker*. Available at http://www.newyorker.com/magazine/2012/07/23/words-on-trial [Last accessed 11 August 2015].

Hollingsworth, C. 2009. *Alice Beyond Wonderland*. Iowa City: University of Iowa Press.

Hunt, P. 1994. *An Introduction to Children's Literature*. Oxford [England]: Oxford University Press.

Inaki, A. & Okita, T. 2005. A Small-Corpus-Based Approach to Alice's Roles. *Literary and Linguistic Computing* , 21 (3), 283-294.

Kincaid, J. R. 1973. Alice's Invasion of Wonderland. *PMLA* 88 (1), 92-99.

Knowles, M. & Malmkjaer, K. 1996. *Language and Control in Children's Literature*. London: Routledge.

Lakoff, R. T. 1993. Lewis Carroll: Subversive Pragmaticist. *PRAG*, 3 (4), 367-385.

Leech, G. N., & Short, M. 2007. *Style in fiction: A linguistic introduction to English fictional prose*. Pearson Education.

Leech, G. N. 2008. *Language in Literature*. Harlow, England: Pearson Longman.

Leech, G. N. 1974. *Semantics*. Harmondsworth: Penguin.

Lucas, A.L. 1999. Enquiring mind, rebillious spirit: Alice and Pinocchio as nonmodel children. *Children's Literature in Education*, 30 (3), 157-169.

MacArthur, F. Embodied Figures of Speech: Problem-Solving in Alice's Dream of Wonderland. *Atlantis* 26 (2) 51-62.

Mahlberg, M., Conklin, K. & Bisson, M.-J. 2014. Reading Dickens's Characters: Employing Psycholinguistic Methods to Investigate the Cognitive Reality of Patterns in Texts. *Language and Literature,* 23 (4), 369-388.

Mahlberg, M. 2012. *Corpus Stylistics and Dickens's Fiction*. London: Routledge.

Mahlberg, M. 2010. Corpus Linguistics and the Study of Nineteenth-Century Fiction. *Journal Of Victorian Culture*, 15 (2), 292-298.

Mahlberg, M. 2007. Clusters, Key Clusters and Local Textual Functions in Dickens. *Corpora*, 2 (1), 1-31.

McIntyre, D. & Busse, B. (Eds). 2010. *Language and Style*. London: Palgrave Macmillan, 32-54.

Nikolajeva, M. 2003. Fairy Tale and Fantasy: from Archaic to Postmodern. *Marvels&Tales: Journal of Fairy-Tale Studies* 17 (1), 138-156.

Nørgaard, N., Busse, B., & Montoro, R. 2010. *Key terms in stylistics*. London: Continuum.

Olohan, M. 2004. *Introducing Corpora in Translation Studies.* London and New York: Routledge.

Rackin, D. 1966. Alice's Journey to the End of Night. *PMLA*, 81 (5), 313-326.

Rayson, P., Berridge D., & Francis, B. 2004. Extending the Cochran rule for the comparison of word frequencies between corpora. *Journées internationales d'Analyse statistique des Données Textuelles*, 7, 1-12.

Semino, E., & Short, M. 2004. *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.

Scott, M. & Tribble, C. 2006. *Textual Patterns*. Amsterdam: John Benjamins Pub.

Scott, M. 2006. In search of a bad reference corpus. *AHRC ICT Methods Network Expert Seminar on Linguistics*. Lancaster University, 1-14.

Scott, M. 2015. WordSmith Tools version 6, Stroud: Lexical Analysis Software [Last accessed 11 August 2015].

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. & Carter, R. 2004. *Trust the Text*. London: Routledge.

Stubbs, M. & Barth, I. 2003. Using recurrent phrases as text-type discriminators. A quantitative method and some findings. *Functions of Language*, 10 (1), 61-104.

Stubbs, M. 2005. Conrad in the computer: Examples of quantitative stylistics methods. *Language and Literature*, 14 (1), 5 – 24.

Stubbs, M. 2015. Conrad, Concordance, Collocation: Heart of Darkness or Light at the End of the Tunnel?. *Uni-Trier.De*. Available at https://www.uni trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-2003-conrad-lecture.pdf [Last accessed 11 August 2015].

Teubert, W. & Čermáková, A. 2007. *Corpus Linguistics*. London: Continuum.

Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: J. Benjamins.

Toolan, M. 2006. Top Keyword Abridgements of Short Stories: A Corpus Linguistic Resource?. *Journal Of Literary Semantics*, 35 (2), 181-194.

Weigand, E. 1998. *Contrastive Lexical Semantics*. Amsterdam: J. Benjamins.

Wynne, M. 2006. Stylistics: Corpus Approaches. In K. Brown (Ed.), *Encyclopeadia of Language & Linguitics.* 2nd Ed, Elsevier, 223-226.

Appendices

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas Set
1	ALICE	397	1,46	1	1		2 171,91	0,0000000000	
2	TURTLE	59	0,22	1	1		314,10	0,0000000000	
3	HATTER	56	0,21	1	0		307,66	0,0000000000	
4	MOCK	56	0,21	1	0		307,66	0,0000000000	
5	GRYPHON	55	0,20	1	0		302,16	0,0000000000	
6	DUCHESS	42	0,15	1	0		230,72	0,0000000000	
7	SHE	553	2,03	1	3 837	0,97	224,51	0,0000000000	
8	DORMOUSE	40	0,15	1	1		210,46	0,0000000000	
9	QUEEN	75	0,28	1	81	0,02	206,78	0,0000000000	
10	SAID	462	1,70	1	3 121	0,79	199,10	0,0000000000	
11	MOUSE	43	0,16	1	14		174,52	0,0000000000	
12	HARE	31	0,11	1	3		150,39	0,0000000000	
13	HERSELF	83	0,31	1	186	0,05	148,23	0,0000000000	
14	TONE	40	0,15	1	19		148,10	0,0000000000	
15	CATERPILLAR	28	0,10	1	1		145,23	0,0000000000	
16	MARCH	34	0,13	1	14		130,67	0,0000000000	
17	CAT	37	0,14	1	29		116,57	0,0000000000	
18	SOUP	18	0,07	1	0		98,87	0,0000000000	
19	JURY	17	0,06	1	1		85,78	0,0000000000	
20	IT	595	2,19	1	5 715	1,44	85,37	0,0000000000	
21	KING	63	0,23	1	199	0,05	83,44	0,0000000000	
22	DINAH	14	0,05	1	0		76,89	0,0000000000	
23	RABBIT	47	0,17	1	125	0,03	73,01	0,0000000000	
24	DODO	13	0,05	1	0		71,40	0,0000000000	
25	ITS	57	0,21	1	202	0,05	66,92	0,0000000000	
26	T	218	0,80	1	1 700	0,43	64,82	0,0000000000	
27	HASTILY	16	0,06	1	7		60,54	0,0000000000	
28	COURT	18	0,07	1	12		60,08	0,0000000000	
29	LARGE	33	0,12	1	75	0,02	58,26	0,0000000000	
30	M	63	0,23	1	297	0,07	51,58	0,0000000000	
31	GLOVES	11	0,04	1	2		49,52	0,0000000000	
32	KNAVE	9	0,03	1	0		49,43	0,0000000000	
33	SERPENT	9	0,03	1	0		49,43	0,0000000000	
34	VERY	144	0,53	1	1 075	0,27	48,20	0,0000000000	
35	REPLIED	29	0,11	1	77	0,02	45,10	0,0000000000	
36	OFF	73	0,27	1	411	0,10	44,94	0,0000000000	
37	FAN	10	0,04	1	2		44,37	0,0000000000	
38	WHITING	8	0,03	1	0		43,94	0,0000000000	
39	THE	1 638	6,02	1	20 161	5,09	43,80	0,0000000000	
40	VOICE	48	0,18	1	211	0,05	43,33	0,0000000000	
41	PIGEON	12	0,04	1	7		41,83	0,0000000000	
42	WITNESS	10	0,04	1	3		41,28	0,0000000000	
43	TRIAL	10	0,04	1	3		41,28	0,0000000000	
44	CATS	13	0,05	1	10		41,23	0,0000000000	
45	MAD	15	0,06	1	17		40,40	0,0000000000	
46	LORY	7	0,03	1	0		38,45	0,0000000000	
47	HEDGEHOG	7	0,03	1	0		38,45	0,0000000000	
48	TARTS	7	0,03	1	0		38,45	0,0000000000	
49	SOO	7	0,03	1	0		38,45	0,0000000000	
50	PEPPER	7	0,03	1	0		38,45	0,0000000000	
51	OOP	7	0,03	1	0		38,45	0,0000000000	
52	TWINKLE	8	0,03	1	1		37,79	0,0000000000	
53	SLATES	8	0,03	1	1		37,79	0,0000000000	
54	BILL	17	0,06	1	29		36,62	0,0000000000	
55	TIMIDLY	9	0,03	1	3		36,33	0,0000000001	
56	FOOTMAN	11	0,04	1	8		35,61	0,0000000002	
57	EXECUTIONER	6	0,02	1	0		32,95	0,0000000065	
58	BREAD-AND-BUTTER	6	0,02	1	0		32,95	0,0000000065	
59	SOLDIERS	10	0,04	1	7		32,82	0,0000000072	
60	REMARKED	10	0,04	1	7		32,82	0,0000000072	
61	ANXIOUSLY	14	0,05	1	21		32,57	0,0000000086	
62	PUPPY	7	0,03	1	1		32,55	0,0000000087	
63	CHESHIRE	7	0,03	1	1		32,55	0,0000000087	
64	ADDED	23	0,08	1	69	0,02	32,01	0,0000000124	
65	YOU	411	1,51	1	4 439	1,12	31,58	0,0000000162	
66	WILLIAM	8	0,03	1	3		31,44	0,0000000176	

67	OFFENDED	10	0,04	1	8		31,25	0,0000000197
68	CONVERSATION	9	0,03	1	6		30,04	0,0000000395
69	MORAL	8	0,03	1	4		29,19	0,0000000626

Figure 4: ACHcorp positive keywords: p-value = 0,0000001000000012, statistical test: log likelihood.

N	Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness	P	Lemmas Set
1	ALICE	397	1,46	1	16		2 937,47	0,0000000000	
2	TURTLE	59	0,22	1	2		438,33	0,0000000000	
3	QUEEN	75	0,28	1	38		436,82	0,0000000000	
4	HATTER	56	0,21	1	0		432,67	0,0000000000	
5	GRYPHON	55	0,20	1	0		424,94	0,0000000000	
6	SAID	462	1,70	1	6 964	0,55	407,30	0,0000000000	
7	MOCK	56	0,21	1	9		380,77	0,0000000000	
8	DORMOUSE	40	0,15	1	0		309,02	0,0000000000	
9	RABBIT	47	0,17	1	12		304,03	0,0000000000	
10	DUCHESS	42	0,15	1	3		302,56	0,0000000000	
11	KING	63	0,23	1	75		299,68	0,0000000000	
12	MOUSE	43	0,16	1	19		256,60	0,0000000000	
13	HARE	31	0,11	1	1		230,63	0,0000000000	
14	CATERPILLAR	28	0,10	1	0		216,30	0,0000000000	
15	CAT	37	0,14	1	29		196,55	0,0000000000	
16	SHE	553	2,03	1	13 548	1,07	183,75	0,0000000000	
17	MARCH	34	0,13	1	33		171,20	0,0000000000	
18	IT	595	2,19	1	16 170	1,28	143,87	0,0000000000	
19	SOUP	18	0,07	1	0		139,05	0,0000000000	
20	T	218	0,80	1	4 452	0,35	111,86	0,0000000000	
21	DINAH	14	0,05	1	0		108,15	0,0000000000	
22	DODO	13	0,05	1	0		100,42	0,0000000000	
23	HERSELF	83	0,31	1	997	0,08	98,22	0,0000000000	
24	VERY	144	0,53	1	2 590	0,20	94,56	0,0000000000	
25	PIGEON	12	0,04	1	1		85,69	0,0000000000	
26	BEGAN	58	0,21	1	583	0,05	83,57	0,0000000000	
27	JURY	17	0,06	1	19		82,33	0,0000000000	
28	TONE	40	0,15	1	269	0,02	82,30	0,0000000000	
29	OFF	73	0,27	1	928	0,07	80,59	0,0000000000	
30	LITTLE	128	0,47	1	2 357	0,19	80,49	0,0000000000	
31	CATS	13	0,05	1	6		76,98	0,0000000000	
32	WENT	83	0,31	1	1 206	0,10	76,64	0,0000000000	
33	THE	1 638	6,02	1	61 839	4,88	69,43	0,0000000000	
34	DOWN	102	0,38	1	1 890	0,15	63,36	0,0000000000	
35	FAN	10	0,04	1	3		63,33	0,0000000000	
36	M	63	0,23	1	876	0,07	61,83	0,0000000000	
37	WHITING	8	0,03	1	0		61,80	0,0000000000	
38	VOICE	48	0,18	1	548	0,04	60,24	0,0000000000	
39	OOP	7	0,03	1	0		54,07	0,0000000000	
40	LOBSTERS	7	0,03	1	0		54,07	0,0000000000	
41	LORY	7	0,03	1	0		54,07	0,0000000000	
42	SOO	7	0,03	1	0		54,07	0,0000000000	
43	HEDGEHOG	7	0,03	1	0		54,07	0,0000000000	
44	CHESHIRE	7	0,03	1	0		54,07	0,0000000000	
45	THING	49	0,18	1	623	0,05	54,06	0,0000000000	
46	KNAVE	9	0,03	1	4		53,64	0,0000000000	
47	FOOTMAN	11	0,04	1	12		53,64	0,0000000000	
48	COURT	18	0,07	1	69		53,27	0,0000000000	
49	SLATES	8	0,03	1	2		51,87	0,0000000000	
50	COOK	13	0,05	1	28		50,39	0,0000000000	
51	SERPENT	9	0,03	1	6		49,58	0,0000000000	
52	TWINKLE	8	0,03	1	3		49,03	0,0000000000	
53	MUSHROOM	8	0,03	1	3		49,03	0,0000000000	
54	LOBSTER	7	0,03	1	1		48,09	0,0000000000	
55	PUPPY	7	0,03	1	1		48,09	0,0000000000	
56	TARTS	7	0,03	1	1		48,09	0,0000000000	
57	ALL	182	0,67	1	4 840	0,38	46,78	0,0000000000	
58	MAJESTY	12	0,04	1	26		46,40	0,0000000000	
59	WOW	6	0,02	1	0		46,35	0,0000000000	
60	CROQUET	6	0,02	1	0		46,35	0,0000000000	
61	LIZARD	6	0,02	1	0		46,35	0,0000000000	

62	SNEEZING	6	0,02	1	0		46,35	0,0000000000
63	CURIOUS	19	0,07	1	110		43,61	0,0000000000
64	RE	38	0,14	1	470	0,04	43,38	0,0000000000
65	DANCE	13	0,05	1	40		43,07	0,0000000000
66	SOLDIERS	10	0,04	1	17		42,37	0,0000000000
67	PEPPER	7	0,03	1	3		41,98	0,0000000000
68	GLOVES	11	0,04	1	25		41,72	0,0000000000
69	EAT	18	0,07	1	103		41,65	0,0000000000
70	SIZE	13	0,05	1	43		41,56	0,0000000000
71	LARGE	33	0,12	1	376	0,03	41,50	0,0000000000
72	EXECUTIONER	6	0,02	1	1		40,65	0,0000000000
73	ANXIOUSLY	14	0,05	1	58		39,67	0,0000000000
74	THOUGHT	74	0,27	1	1 490	0,12	38,99	0,0000000000
75	PIG	8	0,03	1	9		38,67	0,0000000000
76	FLAMINGO	5	0,02	1	0		38,62	0,0000000000
77	HOOKAH	5	0,02	1	0		38,62	0,0000000000
78	AGAIN	83	0,31	1	1 781	0,14	38,08	0,0000000000
79	GAME	13	0,05	1	53		37,18	0,0000000000
80	SEA	13	0,05	1	54		36,78	0,0000000000
81	OFFENDED	10	0,04	1	25		36,43	0,0000000001
82	GARDENERS	8	0,03	1	12		35,38	0,0000000003
83	ITS	57	0,21	1	1 066	0,08	34,76	0,0000000008
84	TIMIDLY	9	0,03	1	20		34,45	0,0000000015
85	POOL	11	0,04	1	38		34,40	0,0000000016
86	QUITE	55	0,20	1	1 018	0,08	34,20	0,0000000020
87	CHORUS	6	0,02	1	4		33,06	0,0000000060
88	ADDED	23	0,08	1	232	0,02	33,00	0,0000000063
89	DOESN	16	0,06	1	109		32,58	0,0000000085
90	BREAD-AND-BUTTER	6	0,02	1	5		31,40	0,0000000181
91	JURY-BOX	4	0,01	1	0		30,90	0,0000000243
92	RABBIT-HOLE	4	0,01	1	0		30,90	0,0000000243
93	GUINEA-PIGS	4	0,01	1	0		30,90	0,0000000243
94	JURORS	4	0,01	1	0		30,90	0,0000000243
95	OOTIFUL	4	0,01	1	0		30,90	0,0000000243
96	PORPOISE	4	0,01	1	0		30,90	0,0000000243
97	PLAYERS	4	0,01	1	0		30,90	0,0000000243
98	MAD	15	0,06	1	101		30,82	0,0000000253
99	BILL	17	0,06	1	134	0,01	30,74	0,0000000265
100	TAIL	9	0,03	1	27		30,18	0,0000000365
101	LESSONS	10	0,04	1	38		29,73	0,0000000467
102	HASTILY	16	0,06	1	123		29,56	0,0000000514
103	BOTTLE	10	0,04	1	39		29,31	0,0000000587
104	EXECUTED	6	0,02	1	7		28,70	0,0000000816
105	ABOUT	94	0,35	1	2 365	0,19	28,61	0,0000000856
106	KNOW	87	0,32	1	2 133	0,17	28,52	0,0000000900
107	HEAD	50	0,18	1	968	0,08	28,52	0,0000000900
108	GETTING	22	0,08	1	244	0,02	28,51	0,0000000901

Figure 5: ACOcorp positive keywords: p-value = 0,0000001000000012, statistical test: log likelihood.

a)

70	WE	34	0,13	1	1 268	0,32	-40,07	0,0000000000
71	MAN	5	0,02	1	646	0,16	-54,57	0,0000000000
72	MARY	4	0,01	1	694	0,18	-64,84	0,0000000000
73	AND	866	3,19	1	16 664	4,20	-71,72	0,0000000000
74	HIM	43	0,16	1	2 375	0,60	-119,99	0,0000000000
75	HIS	96	0,35	1	3 721	0,94	-125,12	0,0000000000
76	HE	125	0,46	1	7 795	1,97	-441,35	0,0000000000

b)

109	OWN	10	0,04	1	1 759	0,14	-28,57	0,0000000876
110	HER	248	0,91	1	15 996	1,26	-29,21	0,0000000620
111	BEEN	38	0,14	1	3 905	0,31	-31,11	0,0000000214
112	MISS	4	0,01	1	1 281	0,10	-31,19	0,0000000205
113	BY	57	0,21	1	5 393	0,43	-36,31	0,0000000001
114	HAVE	80	0,29	1	6 878	0,54	36,82	0,0000000000
115	HAS	7	0,03	1	1 757	0,14	-37,39	0,0000000000
116	OF	511	1,88	1	30 944	2,44	-38,70	0,0000000000
117	WHICH	49	0,18	1	5 772	0,46	-58,21	0,0000000000
118	MAN	5	0,02	1	2 131	0,17	-58,68	0,0000000000

119	ME	68	0,25	1	7 427	0,59	-66,25	0,0000000000
120	FROM	36	0,13	1	5 121	0,40	-66,69	0,0000000000
121	HIM	43	0,16	1	6 725	0,53	-97,52	0,0000000000
122	MY	58	0,21	1	8 051	0,64	-101,96	0,0000000000
123	HIS	96	0,35	1	12 301	0,97	-140,76	0,0000000000
124	HE	125	0,46	1	17 158	1,36	-215,45	0,0000000000

Figure 6: negative keywords in a) ACHcorp and b) ACOcorp: p-value = 0,0000001000000012, statistical test: log likelihood.

	Word	Freq.	%	Texts	%	Lemmas	Set
1	SAID THE MOCK TURTLE	19	0,07	1	100,00		
2	SHE SAID TO HERSELF	16	0,06	1	100,00		
3	A MINUTE OR TWO	11	0,04	1	100,00		
4	SAID THE MARCH HARE	8	0,03	1	100,00		
5	WILL YOU WON'T YOU	8	0,03	1	100,00		
6	SAID ALICE IN A	7	0,03	1	100,00		
7	AS WELL AS SHE	6	0,02	1	100,00		
8	IN A GREAT HURRY	6	0,02	1	100,00		
9	IN A TONE OF	6	0,02	1	100,00		
10	MORAL OF THAT IS	6	0,02	1	100,00		
11	THE MORAL OF THAT	6	0,02	1	100,00		
12	WELL AS SHE COULD	6	0,02	1	100,00		
13	WON'T YOU WILL YOU	6	0,02	1	100,00		
14	YOU WON'T YOU WILL	6	0,02	1	100,00		
15	AND THE MORAL OF	5	0,02	1	100,00		
16	AS SHE SAID THIS	5	0,02	1	100,00		
17	I BEG YOUR PARDON	5	0,02	1	100,00		
18	SAID THE DUCHESS AND	5	0,02	1	100,00		
19	SAID THE KING AND	5	0,02	1	100,00		
20	SHE SAID THIS SHE	5	0,02	1	100,00		
21	THE LITTLE GOLDEN KEY	5	0,02	1	100,00		
22	THE POOR LITTLE THING	5	0,02	1	100,00		
23	A TONE OF GREAT	4	0,02	1	100,00		
24	AND WHEN SHE HAD	4	0,02	1	100,00		
25	AS SHE COULD FOR	4	0,02	1	100,00		
26	AT THE BOTTOM OF	4	0,02	1	100,00		
27	AT THE TOP OF	4	0,02	1	100,00		
28	BEAU OOTIFUL SOO OOP	4	0,02	1	100,00		
29	DID NOT LIKE TO	4	0,02	1	100,00		
30	EVERY NOW AND THEN	4	0,02	1	100,00		
31	FOR A MINUTE OR	4	0,02	1	100,00		
32	I SHOULD LIKE TO	4	0,02	1	100,00		
33	IN A LOW VOICE	4	0,02	1	100,00		
34	IN AN OFFENDED TONE	4	0,02	1	100,00		
35	IS THE SAME THING	4	0,02	1	100,00		
36	LIKE THE LOOK OF	4	0,02	1	100,00		
37	OFF WITH HER HEAD	4	0,02	1	100,00		
38	OFF WITH HIS HEAD	4	0,02	1	100,00		
39	OUT OF ITS MOUTH	4	0,02	1	100,00		
40	PLEASE YOUR MAJESTY SAID	4	0,02	1	100,00		
41	SAID ALICE TO HERSELF	4	0,02	1	100,00		
42	SAID THE WHITE RABBIT	4	0,02	1	100,00		
43	SEEMED TO BE NO	4	0,02	1	100,00		
44	SHE CAME UPON A	4	0,02	1	100,00		
45	SHE DID NOT LIKE	4	0,02	1	100,00		
46	SO SHE WENT ON	4	0,02	1	100,00		
47	THE END OF THE	4	0,02	1	100,00		
48	THE KING AND THE	4	0,02	1	100,00		
49	THE KING SAID TO	4	0,02	1	100,00		
50	THE MARCH HARE SAID	4	0,02	1	100,00		
51	THE MOCK TURTLE SAID	4	0,02	1	100,00		
52	THERE SEEMED TO BE	4	0,02	1	100,00		
53	WOULD NOT COULD NOT	4	0,02	1	100,00		
54	YOU JOIN THE DANCE	4	0,02	1	100,00		
55	YOU WILL YOU WON'T	4	0,02	1	100,00		

Figure 7: 4-word clusters in AW